JEUX STOCHASTIQUES

par

Rida Laraki

Table des matières

1. Introduction

Les jeux stochastiques modélisent l'interaction entre des décideurs pouvant influencer leur environnement. Ces jeux ont d'abord été introduits et étudiés par Loyd Shapley (1953). Depuis, la littérature n'a cessé de croître.

Dans un jeu stochastique, les joueurs font face à des buts potentiellement différents. Ils doivent assurer un bon paiement aujourd'hui tout en maintenant une espérance de paiement élevée pour demain.

Les jeux stochastiques utilisent des outils mathématiques très variés. Nous allons présenter ici quelques résultats classiques, principalement pour les jeux à somme nulle. Plus précisément, les sections 2 à 4 présentent le modèle. Les sections (5, 6 puis 12) sont dédiées aux jeux à n joueurs et somme non nulle. Les sections 7 à 11 sont dédiées aux jeux à somme nulle.

Ce texte s'est inspiré principalement du cours sur les jeux répétés à somme nulle de Sorin (2002), du cours NATO sur les jeux stochastiques et leurs applications édité par Neyman et Sorin (2003), du polycopié de cours de DEA — non publié — sur les jeux stochastiques par Solan (2006) et, enfin, d'un article sur l'étude asymptotique des jeux absorbants à somme nulle par Laraki (2006).

2. Déroulement

Nous considérons un espace d'états (ou d'environnements) Ω fini. Nous avons un ensemble fini de joueurs noté $N = \{1, \ldots, |N|\}$. Dans chaque état ω le joueur i aura un ensemble d'actions (par étape) $A^i(\omega)$ considéré lui aussi fini (et non vide). $A(\omega) = \Pi_i A^i(\omega)$ est donc l'ensemble de tous les profils d'actions admissibles en une étape donnée à l'état ω. Nous notons l'ensemble des couples (état, profil d'actions) par :

$$\Omega A = \{(\omega, a) \mid a \in A(\omega)\}.$$

Donnons nous aussi une famille de probabilités de transition $q : \Omega A \to \Delta(\Omega)$ où $\Delta(X)$ est l'ensemble des probabilités sur X, et un état initial ω_1. Enfin, soit $g^i : \Omega A \to \mathbb{R}$ la fonction de paiement d'étape du joueur i.

Le jeu se déroule comme suit :

– *Étape 1* : l'état initial est noté ω_1. De manière simultanée et indépendante, chaque joueur choisit une action dans son ensemble d'actions admissibles en ω_1. Si le profil $a_1 = (a_1^i)_{i \in N} \in A(\omega)$ a été choisi, chaque joueur i reçoit pour l'étape 1 le paiement $g_1^i = g^i(\omega_1, a_1)$. Un état ω_2 est alors tiré selon la distribution de probabilité $q(\omega_1, a_1)$.

Ensuite, le couple (état, profil d'actions) (a_1, ω_2) est annoncé publiquement à tous les joueurs.

Le déroulement est maintenant défini par induction.

– *Étape* $t \geqslant 2$: connaissant l'histoire passée h_t,

$$h_t = (\omega_1, a_1, \ldots, \omega_{t-1,} a_{t-1}, \omega_t),$$

simultanément, chaque joueur choisit de façon indépendante aux autres une action dans son ensemble d'actions admissibles en ω_t. Si le profil $a_t = (a_t^i)_{i \in N}$ est choisi, chaque joueur i reçoit pour l'étape t le paiement $g_t^i = g^i(\omega_t, a_t)$. Un état ω_{t+1} est alors tiré selon la distribution de probabilité $q(\omega_t, a_t)$. Enfin, le couple (profil d'actions,état) (a_t, ω_{t+1}) est annoncé publiquement à tous les joueurs.

À titre d'exemple, la pêche est une industrie importante au Royaume-Uni et en Islande. Les deux pays partagent le même territoire dans l'Atlantique. Chaque année ils doivent fixer des quotas pour leurs pêcheurs respectifs. La décision des quotas est déterminée chaque année par rapport au nombre moyen z de kg de poisson par km^2. Celui-ci est mesuré chaque année à la fin de septembre. Les pêcheurs pêchent généralement l'intégralité de leurs quotas. Le taux de croissance des poissons est supposé être de 2% $(1 - \exp(-cz))$ où c est une constante fixée. Le gain pour le Royaume-Uni est mesuré par le nombre x de kg de poissons pêchés par ses pêcheurs par km^2. Le gain pour l'Islande est mesuré par le nombre y de kg de poissons pêchés par ses pêcheurs par km^2. En supposant que le Royaume-Uni a un pouvoir de négociation par rapport à l'Islande égal à $\alpha \in [0, 1]$, cette description peut s'écrire comme un jeu stochastique (avec un espace d'état et des ensembles d'actions compacts mais qui peuvent tous être discrétisés). La variable d'état serait alors $\omega = z$. Quand l'état est z, le Royaume-Uni peut choisir x dans l'intervalle $A^1(z) = [0, \alpha z]$ et l'Islande peut choisir y dans l'intervalle $A^1(z) = [0, (1 - \alpha)z]$. Si l'état actuel est z_t et que x_t et y_t ont été sélectionnés l'état demain sera,

$$z_{t+1} = (z_t - x_t - y_t) + (z_t - x_t - y_t) \times 2\% \, (1 - \exp(-c(z_t - x_t - y_t)) \, .$$

La loi de transition est donc déterministe. Enfin la fonction de paiement du Royaume-Uni est $g^1(\omega, x, y) = x$ et celle de l'Islande est $g^2(\omega, x, y) = y$.

3. Stratégies

C'est essentiellement la même définition que dans le texte sur les jeux répétés (ce volume). Ici nous l'adaptons à notre contexte. Pour tout entier (ou étape) t, l'*ensemble de toutes les histoires* possibles jusqu'à la date t est noté :

$$H_t = (\Omega A)^{t-1} \times \Omega.$$

Un élément de H_t sera noté h_t et la dernière composante est notée ω_t. H_1 est identifié avec l'espace d'état Ω. La première histoire n'est autre que ω_1. L'ensemble de toutes *les histoires de longueur finie* est noté

$$H = \bigcup_{t \geqslant 1} H_t.$$

Enfin, l'espace de toutes les *histoires d'une longueur infinie* (appelé *l'ensemble des parties*) est noté :

$$H_\infty = (\Omega A)^{\mathbb{N}^*}.$$

Pour chaque date t, H_t définit une partition (une algèbre ou un cylindre) de H_∞ : à chaque histoire de longueur finie $h_t \in H_t$ est associée l'ensemble des histoires infinies qui coïncident avec h_t jusqu'à l'étape t. Nous notons cette algèbre par $\mathcal{H}_t$ et notons par $\mathcal{H}$ la σ-algèbre générée par tous les cylindres.

Une stratégie de comportement σ^i pour le joueur i est une fonction qui associe à chaque histoire de longueur finie une action mixte dans $\Delta(A^i)$. L'ensemble des stratégies de comportement du joueur i est noté Σ^i.

Une stratégie de comportement est *pure* si pour chaque histoire finie h_t, $\sigma^i(h_t)$ est pure (est une masse de dirac).

Une stratégie *mixte* est une distribution de probabilité sur les stratégies pures.

Une stratégie de comportement est dite *stationnaire* si, pour chaque couple d'histoires de longueur finie $h_t = (\omega_1, a_1, \ldots, \omega_{t-1,} a_{t-1}, \omega_t)$ et $\widehat{h}_{\widehat{t}} = (\widehat{\omega}_1, \widehat{a}_1, \ldots, \widehat{\omega}_{\widehat{t}-1,} \widehat{a}_{\widehat{t}-1}, \widehat{\omega}_{\widehat{t}})$,

$$\omega_t = \widehat{\omega}_{\widehat{t}} \Longrightarrow \sigma(h_t) = \sigma(\widehat{h}_{\widehat{t}}).$$

Une stratégie stationnaire pour le joueur i sera notée x^i et un profil sera noté par $x = (x^1, \ldots, x^{|N|})$. L'ensemble des stratégies stationnaires du joueur i sera noté X^i, qui peut être identifié à

$\prod_{\omega\in\Omega}\Delta\left(A^i(\omega)\right)$. Ainsi, le nombre de stratégies stationnaires pures du joueur i est $\prod_{\omega\in\Omega}\left|A^i(\omega)\right|$ (où $|F|$ désigne le cardinal de l'ensemble F).

Une stratégie est dite *markovienne* si elle dépend seulement de l'état en cours et du nombre d'étapes écoulées. Mathématiquement, une stratégie de comportement est markovienne si, pour chaque couple d'histoires de même longueur $h_t = (\omega_1, a_1, \dots, \omega_{t-1}, a_{t-1}, \omega_t)$ et $\widehat{h}_t = (\widehat{\omega}_1, \widehat{a}_1, \dots, \widehat{\omega}_{t-1}, \widehat{a}_{t-1}, \widehat{\omega}_t)$,

$$\omega_t = \widehat{\omega}_t \Longrightarrow \sigma(h_t) = \sigma(\widehat{h}_t).$$

Chaque profil de stratégies σ et chaque état initial ω_1 définissent une unique distribution de probabilité sur H_∞ (voir le texte sur les jeux répétés (ce volume)). Cette probabilité sera notée $\mathbb{P}_{\omega_1,\sigma}$ et l'espérance mathématique associée sera noté $\mathbb{E}_{\omega_1,\sigma}$.

4. Objectifs

Ce sont essentiellement les mêmes définitions que celles du le texte sur les jeux répétés (ce volume). Ici, nous les adaptons à notre contexte et les reprenons pour préserver une certaine indépendance entre les textes de ce volume.

4.1. Approche compacte. — Dans cette approche on cherche une modélisation qui permet l'application directe des théorèmes standards d'existence de l'équilibre. On suppose donc que chaque joueur i cherche à maximiser une fonction de paiement total. Une telle fonction ϕ^i associe à chaque suite de paiements d'étapes $\left\{g^i(\omega_t, a_t)\right\}_{t=1,\dots}$ une valeur dans l'intervalle $[-M, M]$ que l'on peut interpréter comme le paiement moyen par étape, où

$$M = \max_{(i,\omega,a)\in I\times\Omega A}\left|g^i(\omega, a)\right|.$$

Rappelons que H_∞ est toujours munie de la topologie naturelle induite par les cylindres. On supposera dans l'approche compacte que pour chaque profil de stratégies pures des autres joueurs σ^{-i}, la fonction $\sigma^i \mapsto \phi^i\left(\sigma^i, \sigma^{-i}\right)$ où σ^i parcourt l'ensemble des stratégies pures du joueur i, est continue pour la topologie naturelle induite par les cylindres.

L'ensemble de stratégies pures étant clairement compact pour la topologie naturelle, il est possible d'appliquer un théorème standard et d'en déduire l'existence d'un équilibre en stratégies mixtes. Le théorème de Kuhn (voir le texte sur les jeux répétés (ce volume)), nous permet ensuite de déduire l'existence d'un équilibre en stratégies de comportement.

Pour prouver l'existence d'un équilibre ayant des propriétés plus spécifiques (stationnaire ou markovien) il nous faut plus de structure dans les paiements. Les deux types de jeux qui vont suivre nous intéressent tout particulièrement du fait de leur récursivité et de l'intérêt qu'ils suscitent dans les applications.

Le jeu répété fini Γ_T dure T étapes. Dans un tel jeu, on suppose que le paiement (moyen) du joueur i pour une partie $h_\infty = (\omega_1, a_1, \ldots, \omega_t, a_t, \ldots) \in H_\infty$ est

$$g_T^i(h_\infty) = \sum_{t=1}^{T} \frac{1}{T}\, g^i\left(\omega_t, a_t\right).$$

Dans *le jeu escompté* Γ_λ, avec $\lambda \in]0,1]$, on suppose que le paiement du joueur i est

$$g_\lambda^i(h_\infty) = \lambda \sum_{t=1}^{\infty} (1-\lambda)^{t-1} g^i(\omega_t, a_t).$$

Il est facile de voir que ces deux jeux appartiennent à la famille des jeux compacts et donc admettent des équilibres en stratégies de comportement. Deux questions restent à résoudre : trouver des équilibres particuliers (stationnaires ou markoviens) mais aussi étudier le comportement asymptotique quand les joueurs sont de plus en plus patients ($T \to \infty$ ou $\lambda \to 0$).

Dans le cadre des jeux à deux joueurs et à somme nulle, on notera dans la suite $v_T(\omega_1)$ (resp. $v_\lambda(\omega_1)$) *la valeur* du jeu fini T fois (resp. du jeu λ-escompté).

4.2. Approche uniforme. — Nous allons montrer dans le section Big-Match que l'unique stratégie optimale peut dépendre explicitement de la durée T du jeu (respectivement du taux d'escompte λ). On montrera aussi que la stratégie limite, quand $T \to \infty$ (resp. $\lambda \to 0$) peut converger vers une mauvaise stratégie. Ceci justifie l'approche uniforme, dans laquelle on cherche des stratégies *uniformément* *bonnes* : c'est-à-dire presque optimales pour tout jeu fini de durée T assez grande (resp. tout jeu escompté avec un taux d'escompte λ assez petit).

Dans un jeu à somme nulle, on dit que la valeur uniforme $v(\omega_1)$ existe si pour tout $\varepsilon > 0$ il existe un profil de stratégies de comportement $(\sigma, \tau) \in \Sigma^1 \times \Sigma^2$ et il existe $T(\varepsilon)$ tels que pour tout $T \geqslant T(\varepsilon)$,

$$\text{(1)} \qquad \mathbb{E}_{\omega_1,\sigma,\widetilde{\tau}}\left(\sum_{t=1}^{T} \frac{1}{T} g^1(\omega_t, a_t) \right) \geqslant v(\omega_1) - \varepsilon \qquad \forall\, \widetilde{\tau} \in \Sigma^2,$$

$$\text{(2)} \qquad \mathbb{E}_{\omega_1,\widetilde{\sigma},\tau}\left(\sum_{t=1}^{T} \frac{1}{T} g^2(\omega_t, a_t) \right) \leqslant v(\omega_1) + \varepsilon \qquad \forall\, \widetilde{\sigma} \in \Sigma^1.$$

L'équation (1) (resp. l'équation (2)) sera interprétée : le joueur 1 (resp. 2) peut *garantir uniformément* $v(\omega_1)$.

La somme d'Abel pour une suite bornée de réels $\{z_t\}_{t\in\mathbb{N}^*}$ peut s'écrire comme une combinaison convexe infinie des sommes de Césaro :

$$\lambda \sum_{t=1}^{\infty} (1-\lambda)^{t-1} z_t = \sum_{T=1}^{\infty} T\lambda^2 (1-\lambda)^{T-1} \Big(\frac{1}{T} \sum_{t=1}^{T} z_t \Big).$$

avec

$$\sum_{T=1}^{\infty} T\lambda^2 (1-\lambda)^{T-1} = 1.$$

Ainsi, si la moyenne de Césaro $(\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} z_t)$ existe, alors la moyenne d'Abel $(\lim_{\lambda\to 0} \lambda \sum_{t=1}^{\infty} (1-\lambda)^{t-1} z_t)$ existe aussi.

Ceci implique en particulier que si la valeur uniforme existe, alors pour tout $\varepsilon > 0$, il existe un profil de stratégies de comportement $(\sigma, \tau) \in \Sigma^1 \times \Sigma^2$ et il existe $\lambda(\varepsilon)$ tels que pour tout $0 < \lambda \leqslant \lambda(\varepsilon)$ et tout $(\widetilde{\sigma}, \widetilde{\tau}) \in \Sigma^1 \times \Sigma^2$,

$$\mathbb{E}_{\omega_1,\sigma,\widetilde{\tau}}\left(\lambda \sum_{t=1}^{\infty} (1-\lambda)^{t-1} g^1(\omega_t, a_t) \right) \geqslant v_\infty(\omega_1) - \varepsilon,$$

$$\mathbb{E}_{\omega_1,\widetilde{\sigma},\tau}\left(\lambda \sum_{t=1}^{\infty} (1-\lambda)^{t-1} g^2(\omega_t, a_t) \right) \leqslant v_\infty(\omega_1) + \varepsilon.$$

Donc si la valeur uniforme existe, alors nous avons :

$$v = \lim_{\lambda\to 0} v_\lambda = \lim_{T\to\infty} v_T.$$

Pour la définition d'un équilibre uniforme en somme non nulle, consulter le texte sur les jeux répétés (ce volume). Cependant, même dans les jeux à somme nulle, un équilibre uniforme n'existe pas toujours — par exemple dans le Big Match plus bas — d'où le recours à la valeur dans le cas à somme nulle et à des paiements uniformes (les limites de paiements d'ε-équilibres uniformes).

La définition exacte de l'ensemble des paiements d'équilibres uniformes E_0 est similaire à celle de la valeur uniforme (voir Mertens, Sorin et Zamir, p. 403) : on définit E_ε comme l'ensemble des paiements $v = (v^1, v^2, \dots)$ tels qu'on trouve un profil de stratégies et une date T qui vérifient : dans tout jeu fini à au moins T étapes, d'une part le paiement du joueur i est au moins $v^i - \varepsilon$ et d'autre part en déviant, aucun des joueurs i ne peut gagner plus que $v^i + \varepsilon$. Puis E_0 est l'intersection des E_ε pour $\varepsilon > 0$.

Dans la suite nous allons présenter plusieurs techniques proposées dans la littérature pour montrer l'existence et/ou la caractérisation de $\lim_{\lambda\to 0} v_\lambda$, ainsi que l'égalité entre $\lim_{\lambda\to 0} v_\lambda$ et $\lim_{T\to\infty} v_T$. Enfin, nous développerons la preuve difficile concernant l'existence de la valeur uniforme (valable seulement dans notre cadre où tous les ensembles sont finis et où les joueurs se rappellent de tout le passé, observent les états et les actions passées).

4.3. Approche infinie. — Cette approche consiste à définir pour chaque joueur i, une fonction mesurable et bornée par M, définie directement sur l'ensemble des parties H_∞. Les exemples de l'approche compacte y sont inclus mais nous avons aussi les exemples suivants :

– $\liminf_{T\to\infty} \sum_{t=1}^{T} \frac{1}{T}\, g^i(\omega_t, a_t)$;
– $\limsup_{T\to\infty} \sum_{t=1}^{T} \frac{1}{T}\, g^i(\omega_t, a_t)$;
– $\liminf_{\lambda\to 0} \lambda \sum_{t=1}^{\infty} (1-\lambda)^t g^i(\omega_t, a_t)$;
– $\limsup_{\lambda\to 0} \lambda \sum_{t=1}^{\infty} (1-\lambda)^t g^i(\omega_t, a_t)$.

D'après la relation précédente entre les sommes d'Abel et de Césaro nous déduisons que

$$\lim_{T\to\infty} \inf \frac{1}{T} \sum_{t=1}^{T} z_t \leqslant \lim \inf_{\lambda\to 0} \lambda \sum_{t=1}^{\infty} (1-\lambda)^{t-1} z_t.$$

$\lim_{T\to\infty} \inf \frac{1}{T} \sum_{t=1}^{T} z_t$ est donc l'évaluation la plus pessimiste de la suite des paiements d'étapes des quatre : si un joueur garantit un montant pour cette évaluation, il en est de même pour les trois autres évaluations.

5. Équilibre markovien

Théorème 5.1. — *Tout jeu stochastique fini Γ_T à N joueurs admet un équilibre en stratégies markoviennes.*

Démonstration. — Nous montrons ce résultat par récurrence sur la longueur du jeu T. Pour cela nous utilisons un argument de programmation dynamique.

Sans perte de généralité, nous supposerons que les joueurs maximisent la somme des paiements d'étapes (au lieu de la somme divisée par la longueur du jeu T).

Pour $T = 1$ c'est trivial.

Supposons que le résultat est vrai pour un jeu qui dure T étapes.

Soit le jeu de longueur $T + 1$ et d'état initial ω_1. Quand les joueurs ont joué une fois, on arrive à un état ω_2 et il reste à jouer T étapes. En utilisant l'hypothèse de récurrence, nous sélectionnons pour chaque état futur possible ω_2 un équilibre markovien dans le jeu qui dure T étapes. Chaque joueur décide alors de jouer, à partir de l'étape 2, l'équilibre markovien du jeu fini qui dure T étapes.

Soit $f^i(\omega_2)$ le paiement global d'un tel équilibre (en sommant les paiements d'étapes à partir de l'étape 2). Ainsi, à l'étape 1, on peut considérer que les joueurs font face au jeu statique suivant :

– L'ensemble des joueurs est N ;
– L'ensemble d'actions du joueur i est $A^i(\omega_1)$;
– La fonction de paiement du joueur i est :

$$r^i(\omega_1, a) = g^i(\omega_1, a) + \sum_{\omega_2 \in \Omega} q(\omega_1, a)(\omega_2) f^i(\omega_2).$$

Ce jeu fini admet un équilibre en stratégies mixtes par application du théorème de Nash. Supposons alors que dans le jeu fini de $T + 1$ étapes, les joueurs commencent par jouer un équilibre du jeu statique puis jouent l'équilibre markovien sélectionné dans le jeu fini qui dure T étapes. Ceci définit un équilibre markovien du jeu fini en $T + 1$ étapes. □

6. Équilibre stationnaire

Shapley (1953), l'inventeur du modèle des jeux stochastiques, a montré l'existence de la valeur et des stratégies optimales stationnaires pour les jeux escomptés à deux joueurs et à somme nulle. Nous montrons ici le résultat de Fink (1964) et Takahashi (1964) qui généralisent Shapley aux jeux à n joueurs.

On commence par proposer une méthodologie générale pour calculer explicitement le paiement λ-escompté $g^i_\lambda(\omega, x)$ pour un profil de stratégies stationnaires $x = (x^i)_{i\in N}$.

Proposition 6.1. — $\left\{g^i_\lambda(\omega, x)\right\}_{\omega\in\Omega}$ *est l'unique solution au système linéaire de* $|\Omega|$*-équations :*

$$g^i_\lambda(\omega, x) = \sum_{a\in A(\omega)} \Big(\prod_{i\in N} x^i(a^i)\Big) \begin{pmatrix} \lambda g^i(\omega, a) \\ +\,(1-\lambda)\sum_{\omega'\in\Omega} q(\omega, a)(\omega') g^i_\lambda(\omega', x) \end{pmatrix}.$$

Démonstration. — Puisque les joueurs jouent des stratégies stationnaires, l'espérance de paiement dépend seulement de l'état courant. Donc $\left\{g^i_\lambda(\omega, x)\right\}_{\omega\in\Omega}$ satisfait nécessairement ce système d'équations. Pour montrer que c'est l'unique solution, nous utilisons un principe de maximum. Supposons qu'il y ait deux solutions, $\{\alpha(\omega)\}_{\omega\in\Omega}$ et $\{\beta(\omega)\}_{\omega\in\Omega}$, au système. Soit

$$\omega_0 \in \arg\max_{\omega\in\Omega} \alpha(\omega) - \beta(\omega),$$

et supposons sans perte de généralité que

$$\alpha(\omega_0) - \beta(\omega_0) = k \geqslant 0.$$

Alors,

$$\begin{aligned} k &= \alpha(\omega_0) - \beta(\omega_0) \\ &= \sum_{a\in A(\omega_0)} \Big(\prod_{i\in N} x^i(a^i)\Big)\Big((1-\lambda)\sum_{\omega'\in\Omega} q(\omega_0, a)(\omega')\Big)\big(\alpha(\omega') - \beta(\omega')\big) \\ &\leqslant (1-\lambda)k. \end{aligned}$$

D'où $k = 0$. □

Nous en déduisons les propriétés suivantes.

Proposition 6.2. — *Pour tout joueur i et tout état initial ω, la fonction $(\lambda, x) \mapsto g^i_\lambda(\omega, x)$ est continue et une fraction rationnelle en λ.*

Démonstration. — Puisque le système est linéaire en λ, la solution est nécessairement une fraction rationnelle en λ. La continuité découle du fait que le système linéaire admet une unique solution. □

Nous allons maintenant utiliser la continuité pour montrer que le jeu Γ_λ admet une stationnarité qui permet l'existence d'un équilibre stationnaire. Pour cela, on définit une famille de jeux auxiliaires statiques similaire à celle de la section précédente. Pour chaque $|N|$-uplet de fonctions $f^1, \ldots, f^{|N|}$ de $\Omega \to \mathbb{R}$ bornées par M et chaque $\omega_1 \in \Omega$, on considère un jeu en un coup $G_{\omega_1}(f^1, \ldots, f^{|N|})$ défini comme suit :

– l'ensemble des joueurs est N ;
– l'ensemble d'actions du joueur i est $A^i(\omega)$;
– la fonction de paiement du joueur i est :

$$r^i(\omega_1, a) = \lambda g^i(\omega_1, a) + (1-\lambda) \sum_{\omega_2 \in \Omega} q(\omega_1, a)(\omega_2) f^i(\omega_2).$$

***Théorème 6.3* (Fink (1964), Takahashi (1964)).** — *Tout jeu stochastique escompté à N joueurs admet un équilibre en stratégies stationnaires.*

Démonstration

Étape 1 : nous appliquons le théorème de Kakutani. Considérons l'ensemble

$$XF := \prod_{i \in N} \prod_{\omega \in \Omega} \Delta\left(A^i(\omega)\right) \times [-M, M]^{|N| \times |\Omega|}.$$

C'est un convexe compact d'un espace euclidien de dimension finie.

Un élément de XF sera noté

$$(x, f) = \left(\left\{x^i(\omega)\right\}_{\omega \in \Omega}^{i \in N}, \left\{f^i(\omega)\right\}_{\omega \in \Omega}^{i \in N}\right).$$

$x^i(\omega)$ doit être interprétée comme la stratégie jouée par le joueur i si l'état aujourd'hui est ω et $f^i(\omega)$ comme étant le paiement de continuation du joueur i si le nouvel état est ω.

Nous allons définir une correspondance $W : XF \to XF$ comme suit. Si les coordonnées de W sont notées

$$W = \left(W_X^{i,\omega}, W_F^{i,\omega}\right)_{\omega \in \Omega}^{i \in N},$$

alors

$$\begin{aligned} &W_X^{i,\omega}(x,f) \\ &\quad = \arg \max_{y^i \in \Delta(A^i(\omega))} \left\{ \begin{array}{c} \lambda g^i(\omega, y^i, x^{-i}(\omega)) \\ +(1-\lambda) \sum_{\omega' \in \Omega} q(\omega, y^i, x^{-i}(\omega))(\omega') f^i(\omega') \end{array} \right\}, \end{aligned}$$

et

$$W_F^{i,\omega}(x,f) = \lambda g^i(\omega, x(\omega)) + (1-\lambda) \sum_{\omega' \in \Omega} q(\omega, x)(\omega') f^i(\omega').$$

où la notation (y^i, x^{-i}) veut dire que le joueur i utilise la stratégie stationnaire y^i et les autres joueurs le profil x^{-i}. La motivation est la suivante. Considérons le jeu en un coup $G_\omega(f^1, \ldots, f^{|N|})$. Alors, $W_F^{i,\omega}(x,f)$ est le paiement espéré du joueur i si les joueurs utilisent le profil de stratégies stationnaires x et $W_X^{i,\omega}(x,f)$ est l'ensemble de toutes les meilleures réponses possibles du joueur i face à x^{-i}.

Il est facile de vérifier que cette correspondance satisfait aux hypothèses du théorème de Kakutani (pour la semi-continuité supérieure, on utilise la propriété de continuité dans la proposition 6.2). Nous concluons alors à l'existence d'un point fixe que nous notons (x,f).

Étape 2 : nous montrons que $g_\lambda^i(\omega, x) = f^i(\omega)$.

Ceci résulte de la proposition 6.1. En effet, puisque (x,f) est un point fixe de W, nous avons que pour tout i et tout ω :

$$f^i(\omega) = \lambda g^i(\omega, x(\omega)) + (1-\lambda) \sum_{\omega' \in \Omega} q(\omega, x(\omega))(\omega') f^i(\omega').$$

Étape 3 : nous prouvons que pour toute stratégie σ^i du joueur i, $g_\lambda^i(\omega, \sigma^i, x^{-i}) \leqslant g_\lambda^i(\omega, x)$.

Par définition de F, nous avons que pour tout ω, $x^i(\omega)$ est une meilleure réponse du joueur i contre $x^{-i}(\omega)$ dans le jeu $G_\omega(f^1, \ldots, f^{|N|})$. D'où, pour tout ω et tout $y^i(\omega) \in \Delta\left(A^i(\omega)\right)$,

$$\begin{aligned} \lambda g^i(\omega, y^i, x^{-i}(\omega)) &+ (1-\lambda) \sum_{\omega' \in \Omega} q(\omega, y^i, x^{-i}(\omega))(\omega') f^i(\omega') \\ &\leqslant \lambda g^i(\omega, x(\omega)) + (1-\lambda) \sum_{\omega' \in \Omega} q(\omega, x(\omega))(\omega') f^i(\omega') \\ &= f^i(\omega) \\ &= g_\lambda^i(\omega, x). \end{aligned}$$

Soit $h_t = (\omega_1, a_1, \ldots, \omega_{t-1}, a_{t-1}, \omega_t)$ une histoire partielle de durée t et soit σ^i une stratégie quelconque du joueur i (qui peut dépendre de toute l'histoire du jeu). La dernière inégalité permet alors de déduire que :

$$\mathbb{E}_{\omega_1,\sigma^i,x^{-i}}\left(\begin{matrix}\lambda g^i(\omega_t, \sigma^i(h_t), x^{-i}(\omega_t)) \\ +(1-\lambda)\sum\limits_{\omega_{t+1}\in\Omega} q(\omega_t, \sigma^i(h_t), x^{-i}(\omega_t))(\omega_{t+1}) g^i_\lambda(\omega_{t+1}, x)\end{matrix} \,\middle|\, h_t\right) \leqslant g^i_\lambda(\omega_t, x).$$

Ceci implique, après une sommation, que

$$\begin{aligned} g^i_\lambda(\omega_1, \sigma^i, x^{-i}) &= \mathbb{E}_{\omega_1,\sigma^i,x^{-i}}\left(\lambda\sum_{t=1}^{\infty}(1-\lambda)^{t-1} g^i(\omega_t, a_t)\right) \\ &\leqslant g^i_\lambda(\omega_1, x). \end{aligned}$$

Nous avons donc bien un équilibre de Nash stationnaire. □

7. Opérateur de Shapley

À partir de maintenant et sauf mention explicite, nous nous focaliserons sur les jeux à deux joueurs et à somme nulle.

Puisque l'état du jeu est connu des deux joueurs, chaque joueur peut écrire le principe de programmation dynamique pour calculer sa stratégie optimale. En fait, v_λ et v_T peuvent être calculés à l'aide d'un même opérateur appelé l'opérateur de Shapley. Celui-ci étend le principe de programmation dynamique de Bellman. Ce principe a été publié par Shapley avant et indépendamment de Bellman. De plus Shapley traite le cas de deux joueurs alors que Bellman considère seulement celui d'un seul joueur.

L'opérateur de Shapley n'est autre que l'opérateur valeur pour un jeu statique, similaire à celui introduit dans la preuve du théorème 5.1. Pour chaque état possible ω_1 et chaque fonction $f : \Omega \to \mathbb{R}$ bornée par M, on introduit le jeu statique suivant à deux joueurs et à somme nulle :

– l'ensemble de stratégies pures du joueur 1 est $A^1(\omega_1)$;
– l'ensemble de stratégies pures du joueur 2 est $A^2(\omega_1)$;

– la fonction de paiement du joueur 1 est :

$$r^1(a,\omega_1) = g^1(a,\omega_1) + \sum_{\omega_2\in\Omega} q(\omega_1,a)(\omega_2)f(\omega_2).$$

Ce jeu admet une valeur en stratégies mixtes, noté $\Psi\,(f)\,(\omega_1)$. L'opérateur de Shapley Ψ est défini sur l'espace $\mathcal{F}$ des fonctions $f:\Omega\to\mathbb{R}$ qui sont bornées par M.

Pour tout $\lambda\in\,]0,1[$ nous définissons l'opérateur Φ comme suit :

$$\Phi\,(\lambda,f) = \lambda\Psi\Big(\frac{1-\lambda}{\lambda}f\Big).$$

$\Phi\,(\lambda,f)\,(\omega_1)$ correspond à la valeur en stratégies mixtes du jeu statique ayant comme paiement :

$$\lambda g^1(a,\omega_1) + (1-\lambda)\sum_{\omega_2\in\Omega} q(\omega_1,a)(\omega_2)f(\omega_2).$$

L'analyse de la section précédente montre le résultat suivant.

Théorème 7.1. — *v_λ est l'unique élément de $\mathcal{F}$ qui vérifie*

$$v_\lambda = \Phi\,(\lambda,v_\lambda)\,.$$

La suite $\{v_T\}_{T\in\mathbb{N}^*}$ est l'unique suite dans $\mathcal{F}$ solution de :

$$\begin{aligned} v_{T+1} &= \Phi\Big(\frac{1}{T+1},v_T\Big);\\ v_0 &= 0. \end{aligned}$$

Il est facile de voir que l'opérateur $\Phi\,(\lambda,\cdot)$ est contractant et donc admet un unique point fixe. Nous n'avons donc pas besoin d'utiliser le théorème de point fixe de Kakutani pour l'existence de v_λ ou de v_T mais seulement du théorème du minmax pour montrer que Φ est bien défini, et de l'argument de contraction de Picard.

Remarquons que dans Γ_λ et Γ_T, les joueurs ont une stratégie optimale markovienne et n'ont donc besoin d'observer que l'état courant pour bien jouer. Nous allons voir que, pour bien jouer uniformément, il est nécessaire que les joueurs observent leurs paiements. Sans cette observabilité, la valeur uniforme n'existerait pas. Ceci montre une différence fondamentale et intuitive entre l'approche compacte et l'approche uniforme : pour bien jouer uniformément, un joueur joue d'une manière élaborée.

Nous nous intéressons à l'étude de v_λ et v_T quand les joueurs sont de plus en plus patients ($\lambda \to 0$ et $T \to \infty$) et à la caractérisation de leur limite si possible. Nous allons commencer par un cas (plus) simple : celui des jeux absorbants.

8. Jeux absorbants

Dans cette section nous étudions une classe de jeu, introduite formellement par Kohlberg (1974) et qui va nous servir pour illustrer certains des résultats exposés et expliquer une partie des difficultés qui peuvent être rencontrées dans les jeux stochastiques.

Un état $\omega \in \Omega$ est dit absorbant si, une fois atteint, les joueurs ne peuvent jamais en sortir. Mathématiquement, cela veut dire que pour tout profil $a \in A(\omega)$, on a $q\,(\omega, a)\,(\omega) = 1$. Un jeu est absorbant s'il admet seulement un unique état non absorbant.

Une fois qu'un état absorbant est atteint, le jeu est réduit à un jeu répété à information parfaite (déjà analysé dans le premier texte de ce volume). Nous savons alors que, partant d'un tel état, un équilibre existe. Si nous nous intéressons à l'analyse des équilibres, on peut supposer, sans perte de généralité, qu'une fois qu'un état absorbant est atteint, la suite des paiements est constante et égale à un paiement d'équilibre (que nous avons préalablement sélectionné dans le jeu répété).

En résumé, on va supposer dans toute la suite et sans perte de généralité, qu'à tout état absorbant dans un jeu stochastique est associé un paiement absorbant (un paiement d'étape que les joueurs reçoivent à toutes les étapes suivantes du jeu). Nous supposerons que l'état initial ω_1 d'un jeu absorbant est l'état non absorbant (sinon le jeu serait trivial et sans enjeu). Dès que l'on quitte cet état, le jeu est essentiellement terminé (il n'y a plus de difficulté mathématique liée à l'aspect stochastique). Il n'est donc plus nécessaire de spécifier l'état de départ dans un jeu absorbant.

Ainsi, un jeu absorbant à somme nulle peut être décrit d'une manière compacte comme suit. Il y a deux joueurs, 1 et 2. Le jeu est donné par deux ensembles finis d'actions, A^1 pour le joueur 1 et A^2 pour le joueur 2. Nous avons par ailleurs besoin de deux fonctions de paiements $\widetilde{g} : A^1 \times A^2 \to \mathbb{R}$ et $g^* : A^1 \times A^2 \to \mathbb{R}$. Enfin, nous avons besoin d'une famille de probabilités de transition $p^* : A^1 \times A^2 \to [0,1]$. Le jeu est joué comme suit. À l'étape $t = 1, 2, \ldots$, le joueur 1 choisit une action $a_t^1 \in A^1$ et simultanément le joueur 2 une action $a_t^2 \in A^2$, puis

(i) avec probabilité $p^*\left(a_t^1, a_t^2\right)$ le jeu est absorbé et le joueur 1 reçoit le paiement $g^*\left(a_t^1, a_t^2\right)$ à chacune des étapes restantes.

(ii) avec probabilité $1 - p^*\left(a_t^1, a_t^2\right) = \widetilde{p}\left(a_t^1, a_t^2\right)$ le jeu n'est pas absorbé et le paiement de l'étape t du joueur 1 est $\widetilde{g}\left(a_t^1, a_t^2\right)$.

Exemple d'un jeu d'arrêt. — Soit le jeu absorbant suivant à deux joueurs et à somme nulle.

	C	A
C	0	1^*
A	1^*	0^*

Dans ce jeu, il y a deux paiements absorbants : 1 et 0 (ils sont marqués par $*$). Le paiement absorbant 1^* est atteint si un des profils d'actions (A,C) ou (C,A) est joué (et dans ce cas le joueur gagne 1 par étape, et ce jusqu'à la fin des temps, et le joueur 2 gagne l'opposé, soit -1). Le paiement absorbant 0^* est atteint si (A,A) est joué. Le jeu reste dans l'état non absorbant si et seulement si les joueurs jouent (C,C) (c'est la seule case de la matrice sans $*$).

Dans un *jeu d'arrêt* général chaque joueur a deux options : A (arrêter) ou C (continuer). Dès qu'un joueur choisit A, le jeu est absorbé (s'arrête) et tant que le jeu continue les joueurs reçoivent un paiement de 0 par étape.

Considérons le profil suivant de stratégies stationnaires dans l'exemple du jeu d'arrêt : le joueur 1 joue $(xH, (1-x)B)$ et le joueur 2 joue $(yG, (1-y)D)$. Calculons maintenant le paiement du joueur 1 $g_\lambda(x,y)$ dans le jeu escompté Γ_λ :

$$g_\lambda(x,y) = xy\left(\lambda \times 0 + (1-\lambda)\, g_\lambda(x,y)\right) + \left((1-x)y + (1-y)x\right),$$

d'où

$$g_\lambda(x,y) = \frac{x + y - 2xy}{1 - xy(1-\lambda)}.$$

Dans cet exemple, la valeur $v_\lambda \in [0,1]$ satisfait :

$$\begin{aligned} v_\lambda &= \text{valeur}\left(\begin{array}{c|c|c|} & \text{C} & \text{A} \\ \cline{2-3} \text{C} & (1-\lambda)v_\lambda & 1 \\ \cline{2-3} \text{A} & 1 & 0 \\ \cline{2-3} \end{array}\right) \\ &= \max_{x\in[0,1]} \min_{y\in[0,1]} \left[xy(1-\lambda)v_\lambda + x(1-y) + y(1-x)\right] \\ &= \min_{y\in[0,1]} \max_{x\in[0,1]} \left[xy(1-\lambda)v_\lambda + x(1-y) + y(1-x)\right]. \end{aligned}$$

On vérifie alors que les joueurs n'ont pas de stratégies optimales pures. Si $x_\lambda \in]0,1[$ (resp. $y_\lambda \in]0,1[$) est la stratégie optimale du joueur 1 (resp. du joueur 2) alors, en utilisant le fait que chaque joueur est indifférent entre ces deux actions (les deux ont la même espérance de paiement) nous trouvons que :

$$v_\lambda = x_\lambda(1-\lambda)v_\lambda + (1-x_\lambda) = x_\lambda = y_\lambda.$$

Donc il existe des stratégies optimales uniques, elles sont stationnaires et l'on a :

$$v_\lambda = x_\lambda = y_\lambda = \frac{1-\sqrt{\lambda}}{1-\lambda}.$$

Remarquons que v_λ n'est pas une fraction rationnelle de λ. Ceci est une différence fondamentale avec les jeux à un seul joueur (programmation dynamique). Dans le cas d'un joueur, il existe toujours une stratégie optimale qui est stationnaire et pure pour tout λ. Ceci implique en particulier que v_λ est une fraction rationnelle de λ (car c'est un maximum sur un ensemble fini de fonctions rationnelles en λ) mais aussi qu'il existe un $\lambda_0 > 0$ et une même stratégie pure qui est optimale pour tout $\lambda < \lambda_0$ (une stratégie uniformément bonne, Blackwell (1962)). Enfin, si le jeu d'arrêt se jouait en temps continu (escompté ou non), le joueur 1 pourrait arrêter seul le jeu avec probabilité 1 et garantir ainsi un paiement proche de 1 : il lui suffit de tirer un instant uniformément entre le début du jeu et un temps très proche du début et d'arrêter le jeu à cet instant. Ceci montre une différence fondamentale entre le comportement optimal en temps discret et celui en temps continu. En effet, en temps discret, le joueur 1 ne peut garantir d'arrêter le jeu seul.

Dans cet exemple, $\lim v_\lambda$ existe et est égale à 1. Dans la section suivante nous allons montrer, par l'utilisation d'une approche semi-algébrique, que la valeur asymptotique existe pour tout jeu stochastique à somme nulle. Dans le cas des jeux absorbants, la convergence peut être démontrée plus simplement, en utilisant une approche variationnelle (Laraki 2006). La méthode s'inspire de l'utilisation des solutions de viscosité pour montrer l'existence et la caractérisation de la valeur dans les jeux différentiels à somme nulle et découle d'une méthodologie plus générale (voir Laraki 2001, 2002). Cette approche dans les jeux absorbants permet une caractérisation explicite de $\lim v_\lambda$ comme la valeur d'un jeu.

Nous avons besoin des *notations suivantes spécifiques à cette section* :

– $\mathbb{R}_+^{A^1} = \{z = (z_{a^1})_{a^1 \in A^1} \mid z_{a^1} \in \mathbb{R}_+\}$ est l'orthant positif associé à l'ensemble fini d'actions du joueur 1 A^1 (qu'on peut identifier à l'ensemble des mesures positives sur A^1).

– Pour $z \in \mathbb{R}_+^{A^1}$, $x \in \Delta(A^1)$ et $a^2 \in A^2$,

- x_{a^1} est la probabilité avec laquelle le joueur 1 joue l'action a^1.
- $\widetilde{g}(x, a^2) = \sum_{a^1 \in A^1} x_{a^1}(1 - p^*(a^1, a^2))\widetilde{g}(a^1, a^2)$ est le paiement non absorbant d'étape si le joueur 1 joue x et le joueur 2 joue j ;
- $g^*(z, a^2) = \sum_{a^1 \in A^1} z_{a^1} p^*(a^1, a^2) g^*(a^1, a^1)$ est l'extension linéaire de $\widetilde{g}(\cdot, a^2)$ à $\mathbb{R}_+^{A^1}$;
- $p^*(z, a^2) = \sum_{a^1 \in A^1} z_{a^1} p^*\left(a^1, a^2\right)$ est la probabilité d'absorption étendue linéairement à $\mathbb{R}_+^{A^1}$;
- $\widetilde{p}(x, a^2) = 1 - p^*(x, a^2)$ est la probabilité de continuation.

L'opérateur de Shapley implique que v_λ existe et est l'unique réel dans $[-M, +M]$ qui satisfait

$$v_\lambda = \max_{x \in \Delta(A^1)} \min_{a^2 \in A^2} \left[\lambda \widetilde{g}(x, a^2) + (1 - \lambda)\, \widetilde{p}(x, a^2) v_\lambda + g^*(x, a^2)\right]. \tag{3}$$

Proposition 8.1. — *v_λ converge vers v quand λ tend vers zéro, où v est donné par la formule suivante :*

$$v = \sup_{z \in \mathbb{R}_+^{A^1}} \sup_{x \in \Delta(A^1)} \min_{a^2 \in A^2} \begin{pmatrix} \frac{g^*(x,a^2)}{p^*(x,a^2)} 1_{\{p^*(x,a^2)>0\}} \\ + \frac{\widetilde{g}(x,a^2) + g^*(z,a^2)}{\widetilde{p}(x,a^2) + p^*(z,a^2)} 1_{\{p^*(x,a^2)=0\}} \end{pmatrix}.$$

Démonstration. — Soit w un point d'accumulation de v_λ quand λ tend vers zéro : $w = \lim_{n\to\infty} v_{\lambda_n}$ où $\lambda_n \to 0$.

Étape 1 : nous allons montrer que $w \leqslant v$. L'idée est de considérer une stratégie stationnaire optimale $x\,(\lambda_n)$ pour le joueur 1 puis d'aller à la limite dans l'opérateur de Shapley. Ce procédé est bien connu dans la théorie du contrôle optimal.

De (3), nous déduisons qu'il existe $x\,(\lambda_n) \in \Delta(A^2)$ tel que,

$$v_{\lambda_n} = \min_{a^2\in A^2} \begin{bmatrix} \lambda_n \widetilde{g}(x(\lambda_n), a^2) + (1-\lambda_n)\, \widetilde{p}(x(\lambda_n), a^2) v_{\lambda_n} \\ +g^*(x(\lambda_n), a^2) \end{bmatrix}.$$

D'où, pour tout $a^2 \in A^2$,

$$v_{\lambda_n} \leqslant \frac{\lambda_n \widetilde{g}(x(\lambda_n), a^2) + g^*(x(\lambda_n), a^2)}{\lambda_n \widetilde{p}(x(\lambda_n), a^2) + p^*(x(\lambda_n), a^2)}. \tag{4}$$

Par compacité de $\Delta(A^1)$, et en considérant si nécessaire une sous-suite de $\{\lambda_n\}$, on peut supposer sans perte de généralité que $x\,(\lambda_n) \to x \in \Delta(A^1)$.

Si $p^*(x, a^2) > 0$, alors en faisant tendre λ_n vers zéro, on trouve que :

$$w = \lim v\,(\lambda_n) \leqslant \frac{g^*(x, a^2)}{p^*(x, a^2)}.$$

Supposons maintenant que $p^*(x, a^2) = \sum_{a^1\in A^1} x_{a^1} p^*(a^1, a^2) = 0$, et donc que pour tout a^1 tel que $x_{a^1} > 0$, nous avons $p^*(a^1, a^2) = 0$. Soit alors $z(\lambda_n) = (x_{a^1}\,(\lambda_n)/\lambda_n)_{a^1\in A^1} \in \mathbb{R}_+^{A^1}$. L'équation (4) devient, après la division par λ_n,

$$\begin{aligned} v\,(\lambda_n) &\leqslant \frac{\widetilde{g}(x(\lambda_n), a^2) + g^*(z(\lambda_n), a^2)}{\widetilde{p}(x(\lambda_n), a^2) + p^*(z(\lambda_n), a^2)} \\ &\triangleq r\left(\lambda_n, a^2\right). \end{aligned}$$

Puisque A^2 est fini, en considérant si nécessaire une sous-suite de $\{\lambda_n\}$, on peut supposer que $r\left(\lambda_n, a^2\right)$ converge pour tout a^2. Puisque $\widetilde{p}(x, a^2) = \lim_{n\to\infty} \widetilde{p}(x(\lambda_n), a^2) = 1$ et $\widetilde{g}(x, a^2) = \lim_{n\to\infty} \widetilde{g}(x(\lambda_n), a^2)$, on en déduit que

$$w \leqslant \lim_{n\to\infty} \frac{\widetilde{g}(x, a^2) + g^*(z(\lambda_n), a^2)}{\widetilde{p}(x, a^2) + p^*(z(\lambda_n), a^2)}.$$

Soit $\varepsilon > 0$. Puisque A^2 est fini, on en déduit l'existence d'un entier $N(\varepsilon)$ tel que, pour tout $a^2 \in A^2$,

– si $p^*(x,a^2) > 0$ alors $w \leqslant \dfrac{g^*(x,a^2)}{p^*(x,a^2)}$.

– si $p^*(x,a^2) = 0$ alors $w \leqslant \dfrac{\widetilde{g}(x,a^2) + g^*(z(\lambda_{N(\varepsilon)}),a^2)}{\widetilde{p}(x,a^2) + p^*(z(\lambda_{N(\varepsilon)}),a^2)} + \varepsilon$.

En conséquence, $w \leqslant v$.

Étape 2 : nous montrons que $w \geqslant v$. L'idée est de construire une stratégie du joueur 1 dans le jeu λ_n-escompté qui lui garantit approximativement v.

Rappelons que λ_n converge vers 0, et que $w = \lim v_{\lambda_n}$.

Soit $(z,x) \in \mathbb{R}_+^{A^1} \times \Delta(A^1)$ ε-optimal pour le joueur 1 dans l'expression de v. Supposons que λ_n soit assez petit. Soit

$$A^1(x) := \{a^1 \in A^1 \mid x_{a^1} = 0\}$$

et définissons $x(\lambda_n) \in \Delta(A^1)$ comme suit :

– si $A^1(x) = \varnothing$ alors $x(\lambda_n) = x$.

– si $a^1 \in A^1(x)$ alors $x_{a^1}(\lambda_n) = z_{a^1} \times \lambda_n$.

– si $a^1 \notin A^1(x)$ et $A^1(x) \neq \varnothing$ alors $x_{a^1}(\lambda_n) = x_{a^1} - \dfrac{\sum_{a^1 \in A^1(x)} z_{a^1}}{|A^1(x)|}\lambda_n$.

Ainsi, nous avons $v(\lambda_n) \geqslant r(\lambda_n) - \varepsilon$, où $r(\lambda_n)$ est l'unique réel dans $[-M,M]$ qui satisfait,

$$r(\lambda_n) = \min_{a^2 \in A^2} \begin{bmatrix} \lambda_n \left[\widetilde{g}(x(\lambda_n),a^2)\right] \\ + (1-\lambda_n)\left(\widetilde{p}(x(\lambda_n),a^2)\right) r(\lambda_n) \\ + g^*(x(\lambda_n),a^2) \end{bmatrix}. \tag{5}$$

En effet, $r(\lambda_n)$ est ce que le joueur 1 s'assure de gagner dans le jeu λ_n-escompté s'il joue la stratégie stationnaire $x(\lambda_n)$. Soit $a^2_{\lambda_n} \in A^2$ une stratégie optimale stationnaire pure pour le joueur 2 contre $x(\lambda_n)$ dans le jeu λ_n-escompté (un élément du arg min dans (5)). Puisque la suite $(a^2_{\lambda_n})_{n\geqslant 1}$ appartient à un ensemble fini A^2, il y a un nombre fini de sous-suites de $\{\lambda_n\}$ pour lesquelles $a^2_{\lambda_n}$ est constant pour n grand. Pour chacune de ces sous-suites, (que l'on continuera à appeler λ_n) on peut supposer que $r(\lambda_n)$ converge vers un certain r. En reprenant exactement le calcul de l'étape 1, nous en déduisons que $w \geqslant v - \varepsilon$ pour tout $\varepsilon > 0$. $\square$

Remarquez qu'on aurait pu obtenir une autre formule en considérant dans la preuve une stratégie optimale du joueur 2 (au lieu du joueur 1).

Coulomb (2001), obtient une formule explicite pour le maxmin uniforme des jeux absorbants avec signaux. Si l'on suppose l'observation parfaite des actions et en utilisant le fait que dans ce cas le maxmin uniforme coïncide avec la valeur uniforme et donc avec $\lim v_\lambda$, nous aboutissons à une autre formule pour $\lim v_\lambda$.

Cette analyse variationnelle et la formule qui en résulte peuvent être étendues aux cas où les ensembles d'actions sont compacts et les fonctions $\widetilde{g}$, g^* et p^* continues. Ceci ne peut être fait dans le cadre de Coulomb (2001) qui utilise une approche semi-algébrique. Cependant, Coulomb étudie le cas uniforme qui nécessite des outils plus complexes et des stratégies plus élaborées...

9. Approche semi-algébrique

Ici, nous montrons l'existence de $\lim_{\lambda\to 0} v_\lambda$ pour tout jeu stochastique à somme nulle en utilisant une approche semi-algébrique, initiée par Bewley et Kohlberg (1976 a et b). Cela permet de montrer en particulier que v_λ est à variation bornée. Cette propriété impliquera que $\lim_{T\to\infty} v_T$ existe et est égale à $\lim_{\lambda\to 0} v_\lambda$. Ici on suit Sorin (2002).

Un ensemble dans $\mathbb{R}^m$ est semi-algébrique s'il peut s'écrire comme union finie d'ensembles A_k de la forme :

$$A_k = \{x \in \mathbb{R}^m \mid p_k(x) > 0\} \quad \text{ou} \quad A_k = \{x \in \mathbb{R}^m \mid p_k(x) = 0\}.$$

où p_k est un polynôme de $\mathbb{R}^m$.

En utilisant le fait que v_λ est l'unique point fixe de $\Phi(\lambda, \cdot)$, on en déduit le résultat suivant.

Proposition 9.1. — *L'ensemble des $(\lambda, v_\lambda, x_\lambda)$ tel que λ parcourt $]0, 1]$, v_λ est la valeur du jeu λ-escompté, x^i_λ est une stratégie optimale stationnaire du joueur i est semi-algébrique.*

Ce résultat peut facilement être étendu aux jeux à somme non nulle en remplaçant valeur par paiement d'équilibre et stratégie optimale par profil d'équilibre de Nash (pour plus de détails sur l'approche semi-algébrique, consulter le chapitre 6 par Neyman dans Neyman et Sorin 2003).

En utilisant l'élimination de Tarski-Seidenberg (Benedetti et Risler 1990, théorème 2.21, p. 54) on en déduit qu'il existe une sélection semi-algébrique par rapport à $\lambda \in\,]0, 1[$. Ceci implique l'existence d'un développement en série de Puiseux au voisinage de 0 pour chaque élément $z_\lambda \in \{x_\lambda, y_\lambda, v_\lambda\}$.

***Théorème 9.2* (Bewley and Kohlberg 1976).** — *Étant donné un jeu stochastique (où tout est fini) à somme nulle, pour chaque $z_\lambda \in \{x_\lambda, y_\lambda, v_\lambda\}$, il existe $\lambda_0 > 0$, $k \in \mathbb{N}^*$, $r_n : \Omega \to \mathbb{R}$, $n = 0, \ldots,$ tels que :*

$$z_\lambda(\omega) = \sum_{n=0}^{\infty} r_n(\omega)\, \lambda^{\frac{n}{k}}.$$

pour tout $\lambda \in]0, \lambda_0[$ et tout $\omega \in \Omega$.

Démonstration. — On applique le résultat de Forster 1981, théorème 8.14, p. 58, au cas où les paiements sont uniformément bornés. □

Ainsi dans l'exemple du jeu d'arrêt nous avons :

$$\begin{aligned} v_\lambda = x_\lambda = y_\lambda &= \frac{1-\sqrt{\lambda}}{1-\lambda} \\ &= \left(1-\sqrt{\lambda}\right)\left(1+\lambda+\lambda^2+\cdots\right). \end{aligned}$$

Dans le cadre des jeux avec plus de deux joueurs ou à somme non nulle, on peut montrer l'existence d'une sélection de stratégies et de paiements d'équilibre ayant un développement en série de Puiseux.

La fonction $\lambda \to f_\lambda$ où $f_\lambda \in \mathcal{F}$ et $\lambda \in]0,1[$ sera dite *à variation bornée* si pour toute suite $\{\lambda_n\}_{n\in\mathbb{N}}$ dans $]0,1[$ décroissante vers 0, nous avons :

$$\sum_{n=1}^{\infty} \left\| f_{\lambda_{n+1}} - f_{\lambda_n} \right\|_\infty < \infty.$$

***Corollaire 9.3*.** — *v_λ est une fonction à variation bornée au voisinage de 0 et admet une limite quand $\lambda \to 0$.*

Démonstration. — La convergence est une conséquence du développement de Puiseux. La variation bornée est une conséquence du fait que pour tout $\lambda \in (0, \lambda_0)$ nous avons :

$$\left\| \frac{dv_\lambda}{d\lambda} \right\|_\infty \leqslant C\lambda^{(1-k)/k}. \qquad \square$$

***Corollaire 9.4*.** — *$\lim_{T\to\infty} v_T$ existe et est égale à $\lim_{\lambda\to 0} v_\lambda$.*

Ceci est une conséquence directe d'un résultat général (voir le chapitre 26 par Neyman dans Neyman et Sorin 2003).

Soit Θ un opérateur d'un espace de Banach $\mathcal{Z}$ dans lui-même qu'on supposera non dilatant, *i.e.* pour tout z et z' dans $\mathcal{Z}$, $\|\Theta(z) - \Theta(z')\| \leqslant \|z - z'\|$. Il est facile de voir alors que, pour tout z et z' dans $\mathcal{Z}$,

$$\left\|\lambda\Theta\left(\frac{1-\lambda}{\lambda}z\right) - \lambda\Theta\left(\frac{1-\lambda}{\lambda}z'\right)\right\| \leqslant (1-\lambda)\left\|z - z'\right\|.$$

de sorte que $w \to \lambda\Theta\left(\frac{1-\lambda}{\lambda}w\right)$ admet un point fixe qui est nécessairement unique (par contraction de Picard). Soit w_λ ce point fixe.

Définissons la suite $\{w_T\}_{T=0,1,\ldots}$ dans $\mathcal{Z}$ par récurrence comme suit : $w_0 = 0$ et $w_{T+1} = \frac{1}{T+1}\Theta\left(Tw_T\right)$, celle-ci pouvant s'écrire aussi $w_T = \frac{\Theta^T(0)}{T}$.

***Théorème 9.5* (Neyman 2003).** — *Si w_λ est à variation bornée alors $\lim_{\lambda\to 0} w_\lambda = \lim_{T\to\infty} w_T$.*

Pour montrer le dernier corollaire il suffit de prendre $\Theta = \Psi$ et $\mathcal{Z} = \mathcal{F}$.

Démonstration. — Remarquons que si w_λ est à variation bornée, elle converge nécessairement vers un certain w. Supposons que $\lambda_t = 1/t$. Dans ce cas,

$$\begin{aligned}\left\|w_{t+1} - w_{\lambda_{t+1}}\right\| &= \left\|\frac{1}{t+1}\Theta\left(tw_t\right) - \frac{1}{t+1}\Theta\left(tw_{\lambda_{t+1}}\right)\right\| \\ &\leqslant \frac{t}{t+1}\left\|w_t - w_{\lambda_{t+1}}\right\| \\ &\leqslant \frac{t}{t+1}\left(\left\|w_t - w_{\lambda_t}\right\| + \left\|w_{\lambda_t} - w_{\lambda_{t+1}}\right\|\right),\end{aligned}$$

soit

$$(t+1)\left\|w_{t+1} - w_{\lambda_{t+1}}\right\| \leqslant t\left\|w_t - w_{\lambda_t}\right\| + t\left\|w_{\lambda_t} - w_{\lambda_{t+1}}\right\|.$$

En sommant on trouve

$$(T+1)\left\|w_{T+1} - w_{\lambda_{T+1}}\right\| \leqslant \left\|w_1 - w_{\lambda_1}\right\| + \sum_{t=1}^{T} t\left\|w_{\lambda_t} - w_{\lambda_{t+1}}\right\|.$$

Puisque la somme $\sum_{t=1}^{T}\left\|w_{\lambda_t} - w_{\lambda_{t+1}}\right\|$ est bornée, nous obtenons que $\frac{1}{T+1}\sum_{t=1}^{T} t\left\|w_{\lambda_t} - w_{\lambda_{t+1}}\right\| \to 0$, d'où le résultat. □

10. Big-Match

Avant d'étudier l'approche uniforme dans les jeux à somme nulle en général, nous étudions le Big-Match. C'est le premier exemple résolu explicitement dans le cadre uniforme par Blackwell et Ferguson (1968). C'est le jeu suivant :

$$\begin{array}{c|c|c|} & \overline{0} & \overline{1} \\ \hline \overline{0} & 1 & 0 \\ \hline \overline{1} & 0^* & 1^* \\ \hline \end{array}$$

L'histoire qu'ils ont utilisée pour le présenter est la suivante :

« Chaque jour le joueur 2 choisit un nombre $\overline{0}$ ou $\overline{1}$, et le joueur 1 essaye de prédire le choix du joueur 2, gagnant un point à chaque fois qu'il a raison. Cela continue tant que le joueur 1 prédit $\overline{0}$. Si un jour il prédit $\overline{1}$, tous les choix futurs pour les deux joueurs sont contraints à être les mêmes que le choix de ce jour : si le joueur 1 avait raison ce jour là, il gagne 1 *point chaque jour suivant ; s'il s'est trompé ce jour là, il gagne* 0 *tous les jours suivants. »*

Seul le joueur 1 peut arrêter le jeu et ceci arrive dès qu'il choisit l'option $\overline{1}$. La formule du théorème 7.1 implique que :

$$v_{T+1} = \frac{1}{T+1} \text{ valeur} \left(\begin{array}{c|c|c|} & \overline{0} & \overline{1} \\ \hline \overline{0} & 1+\mathrm{T}v_T & \mathrm{T}v_T \\ \hline \overline{1} & 0 & 1 \\ \hline \end{array} \right).$$

Par récurrence, on montre que $v_T = 1/2$. La stratégie $(\frac{1}{2}\overline{0}, \frac{1}{2}\overline{1})$ du joueur 2 lui garantit une espérance de gain de $1/2$ par étape et donc $1/2$ uniformément. L'unique stratégie optimale du joueur 1 est markovienne et consiste de jouer $\overline{1}$ avec probabilité $x_m = \frac{1}{m+1}$, s'il reste m étapes à jouer (en particulier il n'existe pas de stratégie stationnaire optimale pour le joueur 1). On calcule facilement que $v_\lambda = 1/2$ et que l'unique stratégie optimale du joueur 1 est évidemment stationnaire. Elle consiste à jouer $\overline{1}$ avec probabilité $x_\lambda = \frac{\lambda}{1+\lambda}$. À la limite quand $\lambda \to 0$, le joueur 1 joue $\overline{0}$ avec probabilité 1 à chaque étape. Cette stratégie est mauvaise car elle ne permet d'assurer que zéro par étape (face à cette stratégie, le joueur 2 pourrait jouer l'option $\overline{1}$ à chaque étape).

Nous allons montrer que le joueur 1 peut garantir uniformément tout paiement strictement moins que $1/2$ (ce qui implique que la valeur uniforme de ce jeu existe et est égale à $1/2$). Cependant, il assez facile de montrer qu'il n'a pas de stratégie uniforme qui garantit exactement $1/2$. Pour le montrer il suffit de remarquer que le joueur 1 doit assurer tout le temps un paiement de continuation de $1/2$ ce qui l'oblige à continuer le jeu avec probabilité 1 à chaque étape (en jouant tout le temps l'option $\overline{0}$). Cette stratégie ne lui garantit qu'un paiement de 0. Le Big-Match n'admet donc pas d'équilibre uniforme mais seulement un paiement d'équilibre uniforme.

L'idée de la construction est la suivante : tant que le jeu n'est pas absorbé, si le joueur 1 a obtenu dans le passé un bon paiement (plus de 1 que de 0), il joue avec une probabilité relativement grande la stratégie $\overline{0}$, sinon il augmente sa probabilité de jouer $\overline{1}$.

***Théorème 10.1* (Blackwell et Ferguson 1968).** — *Pour tout entier $K > 0$ le joueur 1 a une stratégie qui garantit uniformément le paiement $K/2(K+1)$.*

Démonstration. — Fixons un entier $K > 0$. Pour tout nombre d'étapes T, définissons trois entiers :

- 0_T le nombre d'étapes avant T où le joueur 2 joue $\overline{0}$;
- 1_T le nombre d'étapes avant T où le joueur 2 joue $\overline{1}$;
- $k_T = 0_T - 1_T$.

On remarque que $0_T + 1_T = T - 1$, ce qui implique :

$$k_T = 2l_T - T - 1.$$

Si le joueur 2 a souvent joué $\overline{0}$, k_T est grand (en particulier positif). Dans ce cas, le joueur 1 augmentera sa probabilité de jouer $\overline{0}$. Dans le cas contraire, k_T est petit (en particulier négatif) et le joueur 1 augmentera sa probabilité de jouer $\overline{1}$.

Nous définissons la stratégie du joueur 1, σ_K, comme suit : à l'étape T, si le jeu n'a pas été absorbé, jouer $\overline{1}$ avec probabilité $1/(k_T + K + 1)^2$. On note que si $k_T = -K$ alors le jeu est absorbé avec probabilité 1. En particulier, tout choix suivant des joueurs n'affecte plus les paiements. Nous allons montrer que pour toute suite d'actions $a^2 = \left(a_1^2, \ldots, a_t^2, \ldots\right)$ du joueur 2, nous avons :

$$\mathbb{E}_{\sigma_K^1, a^2}\left(\sum_{t=1}^{T} \frac{1}{T}\, g^1\left(\omega_t, a_t\right)\right) \geqslant \frac{K}{2(K+1)} - \frac{K+1}{2T}.$$

Ceci impliquerait que la même propriété est vraie pour toute stratégie pure du joueur 2 et donc aussi pour toute stratégie mixte du joueur 2. Nous déduirons alors que le joueur 1 peut garantir uniformément $\frac{K}{2(K+1)} - \varepsilon$ pour tout ε.

Fixons donc une suite d'actions $a^2 = \left(a_1^2, \ldots, a_t^2, \ldots\right)$ pour le joueur 2. Remarquons alors que :

– si $a_1^2 = \overline{1}$, alors σ_K^1 coïncide avec σ_{K-1}^1 à partir de l'étape 2,
– si $a_1^2 = \overline{0}$, alors σ_K^1 coïncide avec σ_{K+1}^1 à partir de l'étape 2.

Soit t_* la première étape où le jeu est absorbé (le joueur 1 joue $\overline{1}$). Définissons la variable aléatoire X_T comme suit :

$$X_T = \begin{cases} 1/2 & \text{si } t_* > T \\ 1 & \text{si } t_* \leqslant T, \quad a_{t_*}^2 = \overline{1} \\ 0 & \text{si } t_* \leqslant T, \quad a_{t_*}^2 = \overline{0} \end{cases}$$

Si le jeu s'est arrêté avant ou à l'étape T, X_T représente le paiement d'absorption du joueur 1. Sinon, X_T vaut $1/2$ (l'objectif approximatif du joueur 1). Nous allons montrer par récurrence sur T que

$$\mathbb{E}_{\sigma_K^1, a^2}\left(X_T\right) \geqslant \frac{K}{2(K+1)}.$$

Pour $T = 1$:

– Si $a_1^2 = \overline{1}$, alors

$$\begin{aligned} \mathbb{E}_{\sigma_K^1, a^2}\left(X_T\right) &= \left(1 - \frac{1}{(K+1)^2}\right)\frac{1}{2} + \frac{1}{(K+1)^2} \\ &> \frac{1}{2} \\ &> \frac{K}{2(K+1)}. \end{aligned}$$

– Si $a_1^2 = \overline{0}$, alors

$$\begin{aligned}\mathbb{E}_{\sigma^1_K, a^2}(X_T) &= \left(1 - \frac{1}{(K+1)^2}\right)\frac{1}{2} \\ &= \frac{K(K+2)}{2(K+1)^2} \\ &> \frac{K}{2(K+1)}.\end{aligned}$$

Supposons maintenant que la dernière inégalité est satisfaite pour $T = t_0$ et montrons qu'elle est vraie pour $T = t_0 + 1$.

– Si $a_1^2 = \overline{1}$, alors

$$\begin{aligned}\mathbb{E}_{\sigma^1_K, a^2}(X_{t_0+1}) &= \left(1 - \frac{1}{(K+1)^2}\right) E_{\sigma^1_{K-1}, a^2}(X_{t_0}) + \frac{1}{(K+1)^2} \\ &> \left(1 - \frac{1}{(K+1)^2}\right)\frac{K-1}{2K} + \frac{1}{(K+1)^2} \\ &= \frac{K}{2\,(K+1)}.\end{aligned}$$

– Si $a_1^2 = \overline{0}$, alors

$$\begin{aligned}\mathbb{E}_{\sigma^1_K, a^2}(X_{t_0+1}) &= \left(1 - \frac{1}{(K+1)^2}\right) E_{\sigma^1_{K+1}, a^2}(X_{t_0}) \\ &> \left(1 - \frac{1}{(K+1)^2}\right)\frac{K+1}{2(K+2)} \\ &= \frac{K}{2(K+1)}.\end{aligned}$$

Soit $t^* := \min\{t_*, T\}$: c'est l'étape courante si le jeu n'a pas encore été absorbé, sinon c'est l'étape d'absorption. Rappelons que si $k_T = -K$, alors le jeu est absorbé avec probabilité 1. Donc $k_T \geqslant -K$ d'où

$0_T \geqslant (T-K-1)/2$, ce qui implique que :

$$\begin{aligned}
\mathbb{E}_{\sigma_K^1,a^2}\left(\sum_{t=1}^{T}\frac{1}{T}g^1(\omega_t,a_t)\right) &= \mathbb{E}_{\sigma_K^1,a^2}\left(\frac{0_{t^*}+(T-t^*)\,1_{a^2_{t^*}=\bar{1}}}{T}\right)\\
&\geqslant \mathbb{E}_{\sigma_K^1,a^2}\left(\frac{\frac{t^*-K-1}{2}+(T-t^*)\,1_{a^2_{t^*}=\bar{1}}}{T}\right)\\
&= \mathbb{E}_{\sigma_K^1,a^2}\left(\frac{\frac{t^*}{2}+(T-t^*)\,1_{a^2_{t^*}=\bar{1}}}{T}\right)-\frac{K+1}{2T}\\
&= \frac{1}{T}\mathbb{E}_{\sigma_K^1,a^2}\left(\sum_{t=1}^{T}X_t\right)-\frac{K+1}{2T}\\
&\geqslant \frac{K}{2(K+1)}-\frac{K+1}{2T}. \qquad \square
\end{aligned}$$

La stratégie utilisée dans la preuve nécessite l'observation par les joueurs des paiements et de l'état. Sans l'observation des paiements, la valeur uniforme du Big-Match n'existe pas (Coulomb (1992)). Cependant, dans le cas d'un joueur qui n'observe rien (ni état, ni paiements), il peut être démontré qu'une stratégie uniforme existe toujours (Rosenberg, Solan et Vieille (2002)) mais elle n'est pas aussi simple (stationnaire) que la stratégie uniforme de Blackwell (1962) en observation parfaite de l'état.

11. Valeur uniforme

Ici nous étendons le résultat précédent, à savoir l'existence de la valeur uniforme, à tout jeu stochastique à somme nulle.

***Théorème 11.1* (Mertens et Neyman (1981)).** — *Tout jeu stochastique à deux joueurs et à somme nulle admet une valeur uniforme.*

Démonstration. — Comme cela a déjà été montré précédemment, si la valeur uniforme existe, elle est nécessairement égale à $\lim_{\lambda\to 0} v_\lambda(\omega_1) := v(\omega_1)$. Nous allons montrer que pour tout $\varepsilon > 0$, le joueur 1 peut s'assurer uniformément $v(\omega_1)-\varepsilon$. Par symétrie, le joueur 2 peut aussi garantir uniformément $v(\omega_1)+\varepsilon$, d'où le résultat.

Par la section sur l'opérateur de Shapley, nous avons que pour tout taux d'escompte λ, il existe une stratégie optimale stationnaire x^1_λ

pour le joueur 1 (le joueur qui maximise). Donc pour tout $\omega \in \Omega$ et tout $y^2_\lambda \in \Delta\left(A^2(\omega)\right)$ nous avons

$$\lambda g^1(\omega, x^1_\lambda(\omega), y^2_\lambda(\omega)) \\ + (1-\lambda) \sum_{\omega' \in \Omega} q(\omega, x^1_\lambda(\omega), y^2_\lambda(\omega))(\omega') v_\lambda(\omega') \geqslant v_\lambda(\omega).$$

Nous allons construire explicitement une stratégie uniforme ε-optimale pour le joueur 1 comme suit. À chaque étape t, le joueur 1 va jouer la stratégie $x^1_{\lambda_t}(\omega_t)$, où λ_t est une fonction du paiement reçu dans le passé et ω_t est l'état courant. Si le paiement moyen passé est bon, le joueur 1 va augmenter le poids des opportunités futures ($\lambda_{t+1} < \lambda_t$). Dans le cas contraire, il doit essayer d'assurer un bon paiement aujourd'hui ($\lambda_{t+1} > \lambda_t$).

En effet, le taux d'escompte (comme le taux d'intérêt) mesure l'importance du paiement de court terme relativement au paiement de long terme. Dans l'étude de la valeur uniforme, le taux d'escompte peut être arbitrairement petit de telle sorte que les opportunités futures ont un poids très élevé relativement aux opportunités présentes. Cela dit, si le joueur a tout le temps des paiements faibles, il ne va jamais pouvoir utiliser les opportunités futures. Le mécanisme d'adaptation du taux d'escompte dynamiquement en fonction des performances passées permet au joueur de faire proprement son arbitrage entre le paiement aujourd'hui et le paiement demain.

D'après le corollaire 9.3, il existe λ_0 tel que pour tout $0 < \lambda < \lambda' \leqslant \lambda_0$, il existe une fonction positive et intégrable ψ sur $[0, \lambda_0]$ telle que

$$\max_{\omega \in \Omega} \left| v_\lambda(\omega) - v_{\lambda'}(\omega) \right| = \int_\lambda^{\lambda'} \psi(s) ds.$$

Soit $\varepsilon \in \,]0, M[$ fixé. On définit une fonction $D : [0, \lambda_0] \to \mathbb{R}$ comme suit :

$$D(y) = \frac{12M}{\varepsilon} \int_y^{\lambda_0} \frac{\psi(s)}{s} ds + \frac{1}{\sqrt{y}}$$

Nous allons commencer par montrer quelques propriétés sur D avant de prouver le théorème.

Propriété 1. — *D est décroissante et intégrable ($\int_0^{\lambda_0} D(y)dy < +\infty$).*

Démonstration. — La décroissance est évidente. Le second terme, $\frac{1}{\sqrt{y}}$ est clairement intégrable. Pour le deuxième terme, on intervertit les intégrales ce qui permet d'obtenir :

$$\begin{aligned}\int_{y=0}^{\lambda_0}\int_{s=y}^{\lambda_0}\frac{\psi(s)}{s}dsdy &= \int_{s=0}^{\lambda_0}\int_{y=s}^{\lambda_0}\frac{\psi(s)}{s}dsdy \\ &= \int_{s=0}^{\lambda_0}\psi(s)ds \leqslant 2M. \qquad \square\end{aligned}$$

Propriété 2

$$\lim_{y\to 0} D\left(y-\frac{\varepsilon}{6M}y\right)-D(y)=\lim_{y\to 0} D\left(y\right)-D\left(y+\frac{\varepsilon}{6M}y\right)=+\infty.$$

Démonstration. — Nous avons :

$$\lim_{y\to 0} D\left(y-\frac{\varepsilon}{6M}y\right)-D(y)=\lim_{y\to 0}\int_{y-\frac{\varepsilon}{6M}y}^{y}\frac{\psi(s)}{s}ds+\frac{1}{\sqrt{y-\frac{\varepsilon}{6M}y}}-\frac{1}{\sqrt{y}}.$$

Le premier terme est positif et le second tend vers $+\infty$. L'autre limite se démontre de la même façon. $\square$

Nous définissons maintenant une deuxième fonction φ sur $[0,\lambda_0]$ comme suit :

$$\varphi(\lambda)=\int_{D(\lambda)}^{+\infty}D^{-1}(y)dy=\int_0^{\lambda}D(y)dy-\lambda D(\lambda)$$

L'égalité plus haut se déduit géométriquement.

Définissons aussi pour tout entier t la variable aléatoire Z_t comme suit :

$$Z_t = v_{\lambda_t}(\omega_t)-\varphi(\lambda_t).$$

Puisque D est intégrable, quand $\lambda_t\to 0$ la quantité $\varphi(\lambda_t)$ approche 0 et donc Z_t approche $v_0(\omega_t)$. Notre but est de construire une stratégie du joueur 1 telle que Z_t soit une sous-martingale. Puisque la stratégie du joueur 1 va être de jouer la stratégie optimale du jeu λ_t-escompté, il nous faut d'abord construire les variables aléatoires λ_t. Nous allons les construire inductivement.

Nous choisissons λ_1 assez petit pour que

(i) $\varphi(\lambda_1)<\varepsilon$;
(ii) $\forall\, y>D(\lambda_1)$, $D\left(y-\frac{\varepsilon}{6M}y\right)-D(y)>6M$;
(iii) $\forall\, y>D(\lambda_1)$, $D\left(y\right)-D(y+\frac{\varepsilon}{6M}y)>6M$.

Ceci est possible grâce aux propriétés 1 et 2. Nous définissons $\lambda_{t+1} = D^{-1}(d_{t+1})$ où la suite $\{d_t\}$ est définie comme suit :

$$\begin{aligned} d_1 &= D(\lambda_1) \\ d_{t+1} &= \max\left\{D(\lambda_1), d_t + g^1(\omega_t, a_t) - v_{\lambda_t}(\omega_{t+1}) + 4\varepsilon\right\}. \end{aligned}$$

Puisque $d_t \geqslant D(\lambda_1)$ pour $t \geqslant 1$, nous concluons que $\lambda_t \leqslant \lambda_1$.

Nous allons commencer par étudier quelques propriétés utiles des deux suites.

Propriété 3. — $|d_{t+1} - d_t| \leqslant 6M$.

Démonstration. — Cela découle de sa définition puisque $\left|g^1(\omega_t, a_t)\right|$, $|v_{\lambda_t}(\omega_{t+1})|$ et ε sont bornés par M. □

Propriété 4. — $|\lambda_{t+1} - \lambda_t| \leqslant \frac{\varepsilon\lambda_t}{6M}$.

Démonstration. — Étudions le cas $\lambda_{t+1} \leqslant \lambda_t$ (l'autre cas se montre de la même façon). Donc $d_t \leqslant d_{t+1}$ ce qui implique d'après la propriété 3 que $d_{t+1} \leqslant d_t + 6M$.

Supposons maintenant que $\lambda_{t+1} < \lambda_t - \frac{\varepsilon\lambda_t}{6M}$. En utilisant (ii) nous aurons alors que

$$d_{t+1} - d_t \geqslant D(\lambda_t - \frac{\varepsilon\lambda_t}{6M}) - D(\lambda_t) > 6M,$$

ce qui est une contradiction. □

Nous définissons maintenant la stratégie du joueur 1 comme suit : à l'étape t, connaissant l'état actuel ω_t et observant le paiement de l'étape précédente, le joueur actualise le taux d'escompte à λ_t puis à l'étape t il joue la stratégie optimale $x^1_{\lambda_t}(\omega_t)$ dans le jeu λ_t escompté. Appelons cette stratégie σ^1.

On définit

$$\begin{aligned} C_1 &= \varphi(\lambda_t) - \varphi(\lambda_{t+1}); \\ C_2 &= v_{\lambda_{t+1}}(\omega_{t+1}) - v_{\lambda_t}(\omega_{t+1}); \\ C_3 &= \lambda_t\left(g^1\left(\omega_t, a_t\right) - v_{\lambda_t}(\omega_{t+1})\right). \end{aligned}$$

Rappelons que $Z_t := v_{\lambda_t}(\omega_t) - \varphi(\lambda_t)$ et que

$$\varphi(\lambda) = \int_0^\lambda D(y)dy - \lambda D(\lambda).$$

Propriété 5. — *Pour tout t, nous avons :*

$$\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[Z_{t+1} - Z_t\,|\mathcal{H}_t\right] \geqslant \mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[C_1 + C_2 - C_3\,|\mathcal{H}_t\right].$$

Démonstration. — Nous avons, par définition de σ^1

$$\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[\lambda_t g^1(\omega_t, a_t) + (1-\lambda_t) v_{\lambda_{t+1}}(\omega_{t+1})\,|\mathcal{H}_t\right] \geqslant v_{\lambda_t}(\omega_t),$$

d'où

$$\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\begin{bmatrix}\lambda_t\left(g^1(\omega_t, a_t) - v_{\lambda_t}(\omega_t)\right) \\ +\,(1-\lambda_t)\left(v_{\lambda_{t+1}}(\omega_{t+1}) - v_{\lambda_t}(\omega_t)\right)|\mathcal{H}_t\end{bmatrix} \geqslant 0.$$

Donc

$$\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\begin{bmatrix}\lambda_t\left(g^1(\omega_t, a_t) - v_{\lambda_t}(\omega_t)\right) \\ +\,(1-\lambda_t)\left(v_{\lambda_{t+1}}(\omega_{t+1}) - v_{\lambda_t}(\omega_t)\right) \\ +\,v_{\lambda_{t+1}}(\omega_{t+1}) - v_{\lambda_{t+1}}(\omega_{t+1}) \\ +\,\varphi(\lambda_t) - \varphi(\lambda_t) \\ +\,\varphi(\lambda_{t+1}) - \varphi(\lambda_{t+1})\,|\mathcal{H}_t\end{bmatrix} \geqslant 0.$$

Nous en déduisons donc que

$$\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[C_3 - C_2 - C_1 + Z_{t+1} - Z_t\,|\mathcal{H}_t\right] \geqslant 0. \qquad \square$$

Nous allons maintenant borner les termes C_1, C_2 et C_3.

Propriété 6. — $C_1 \geqslant \lambda_t\left(d_{t+1} - d_t\right) - \varepsilon\lambda_t$.

Démonstration. — Considérons le cas $\lambda_{t+1} < \lambda_t$ (l'autre cas se traite de la même façon). Dans ce cas on observe géométriquement, en utilisant la définition de φ et la propriété 1, que

$$\varphi(\lambda_t) - \varphi(\lambda_{t+1}) \geqslant \lambda_t\left(d_{t+1} - d_t\right) - \left(\lambda_{t+1} - \lambda_t\right)\left(d_{t+1} - d_t\right),$$

puis nous utilisons les propriétés 3 et 4. $\square$

Propriété 7. — $|C_2| \leqslant \varepsilon\lambda_t$.

Démonstration. — Nous avons :

$$\begin{aligned}|C_2| &= \left|v_{\lambda_{t+1}}(\omega_{t+1}) - v_{\lambda_t}(\omega_{t+1})\right| \\ &\leqslant \left|\int_{\lambda_t}^{\lambda_{t+1}} \psi(y)dy\right|.\end{aligned}$$

La propriété 4 implique alors que pour tout y entre λ_t et λ_{t+1} nous avons,

$$\frac{\lambda_t}{y} \geqslant \frac{1}{2},$$

d'où

$$\begin{aligned}|C_2| &\leqslant \left|\int_{\lambda_t}^{\lambda_{t+1}} \frac{\psi(y)}{y} dy\right| 2\lambda_t \\ &= \lambda_t \left|\frac{\varepsilon}{6M}(d_t - d_{t+1}) + \frac{1}{\sqrt{\lambda_t}} - \frac{1}{\sqrt{\lambda_{t+1}}}\right|.\end{aligned}$$

On observe par ailleurs que l'intégrale $\int_{\lambda_t}^{\lambda_{t+1}} \frac{\psi(y)}{y} dy$ est positive si et seulement si $\lambda_t < \lambda_{t+1}$. Ceci est vrai si et seulement si $\frac{1}{\sqrt{\lambda_t}} > \frac{1}{\sqrt{\lambda_{t+1}}}$. Nous déduisons donc que $|C_2| \leqslant \frac{\varepsilon}{6M} |d_t - d_{t+1}| \lambda_t$. Il nous suffit donc d'appliquer la propriété 3. □

Propriété 8. — *$C_3 \leqslant \lambda_t (d_{t+1} - d_t) - 4\varepsilon\lambda_t$.*

Démonstration. — Par définition de d_{t+1} nous avons

$$d_{t+1} - d_t \geqslant g^1(\omega, a_t) - v_{\lambda_t}(\omega_{t+1}) + 4\varepsilon,$$

d'où l'inégalité souhaitée. □

Nous en déduisons maintenant la propriété souhaitée.

Propriété 9. — *Z_t est une sous-martingale qui satisfait*

$$\begin{aligned}\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}[Z_t] &\geqslant 2\varepsilon \mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[\sum_{t=1}^{T} \lambda_t\right] + Z_1 \\ &\geqslant Z_1.\end{aligned}$$

Démonstration. — En utilisant les propriétés 5 à 8 nous obtenons que

$$\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}[Z_{t+1} - Z_t \,|\mathcal{H}_t] \geqslant 2\varepsilon\lambda_t.$$

Puis en sommant nous en déduisons l'inégalité recherchée. □

Nous finissons maintenant la preuve du théorème. Pour cela nous allons montrer que pour toute stratégie σ^2 du joueur 2 nous avons

$$\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left(\sum_{t=1}^{T} \frac{1}{T} g^1(\omega_t, a_t)\right) \geqslant v_0(\omega_1) - 7\varepsilon,$$

pour tout $T \geqslant T_0$ que nous choisirons plus tard.

En remplaçant Z_t par sa valeur $v_{\lambda_t}(\omega_t) - \varphi(\lambda_t)$ dans la propriété 9 (la dernière inégalité) nous obtenons que

$$\begin{aligned}\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[v_{\lambda_t}(\omega_t)\right] &\geqslant v_{\lambda_1}(\omega_1) + \varphi(\lambda_t) - \varphi(\lambda_1)\\ &\geqslant v_{\lambda_1}(\omega_1) - \varphi(\lambda_1)\\ &\geqslant v_{\lambda_1}(\omega_1) - \varepsilon,\end{aligned}$$

où la dernière inégalité utilise (i). Nous en déduisons donc que l'espérance du paiement λ_t-escompté en ω_t est élevée.

Par (i) et par définition de Z_t nous obtenons que $|Z_t| \leqslant 2M$. Ainsi, utilisant la propriété 9, nous en déduisons que

$$\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[\sum_{t=1}^{+\infty}\lambda_t\right] \leqslant \frac{M}{\varepsilon},$$

d'où

$$\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[\sum_{t=1}^{+\infty}\mathbf{1}_{\{\lambda_t=\lambda_1\}}\right] \leqslant \frac{M}{\varepsilon}.$$

Par définition de d_{t+1} et en utilisant la propriété 3 nous obtenons :

$$d_{t+1} - d_t \leqslant g^1(\omega_t, a_t) - v_{\lambda_t}(\omega_{t+1}) + 4\varepsilon + 6M\mathbf{1}_{\{\lambda_{t+1}=\lambda_1\}},$$

en sommant nous obtenons :

$$\begin{aligned}\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[\frac{1}{T}\sum_{t=1}^{T} g^1(\omega_t, a_t)\right] &\geqslant \mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[\frac{1}{T}\sum_{t=1}^{T} v_{\lambda_t}(\omega_{t+1})\right]\\ &\quad + \frac{1}{T}\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\Big[d_{T+1} - d_1\Big] - 4\varepsilon\\ &\quad - \frac{6M}{T}\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[\sum_{t=1}^{+\infty}\mathbf{1}_{\{\lambda_t=\lambda_1\}}\right]\\ &\geqslant v_{\lambda_1}(\omega_1) - \varepsilon - \frac{d_1}{T} - 4\varepsilon - \frac{6M^2}{T\varepsilon},\end{aligned}$$

car $d_{T+1} \geqslant 0$. Il nous suffit maintenant pour conclure de choisir $T \geqslant T_0$ avec

$$T_0 = \max\left\{\frac{d_1}{\varepsilon}, \frac{6M^2}{\varepsilon^2}\right\}. \qquad \square$$

Remarque. — Nous avons prouvé par la même occasion que

$$\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[\liminf \frac{1}{T}\sum_{t=1}^{T} g^1(\omega_t, a_t)\right] \geqslant v_{\lambda_1}(\omega_1) - 7\varepsilon.$$

ce qui montre que tous les exemples de l'approche infinie ont une valeur qui coïncide avec la valeur uniforme.

12. Paris Match

Ce jeu a été introduit et étudié explicitement par Sorin (1986) et il appartient à la famille du Big-Match. L'exemple est célèbre car il montre que dans les jeux stochastiques à somme non nulle, il n'y a aucun lien entre l'approche asymptotique (l'ensemble des paiements d'équilibres de Nash quand le taux d'escompte tend vers zéro) et l'approche uniforme (l'ensemble des paiements d'équilibres de Nash uniformes). Ceci montre aussi la différence fondamentale entre les jeux répétés sans aléas (section 3 du le texte sur les jeux répétés (ce volume)) et les jeux stochastiques. L'exemple de Sorin (1986) est le suivant :

	G	D
H	(1,0)	(0,1)
B	(0,2)*	(1,0)*

Ainsi, en gardant seulement les paiements du joueur 1, nous retrouvons le Big-Match étudié par Blackwell et Ferguson (1968).

Soit E_λ l'ensemble des paiements d'équilibres de Nash du Paris-Match λ-escompté, E_T l'ensemble des paiements d'équilibres de Nash du Paris-Match répété T fois et enfin E_0 l'ensemble des paiements d'équilibres uniformes.

***Théorème 12.1* (Sorin (1986)).** — *E_λ et E_T sont disjoints de E_0. Plus précisément, pour tout λ et T nous avons :*

(a) $E_\lambda = \{(\frac{1}{2}, \frac{2}{3})\}$;
(b) $E_T = \{(\frac{1}{2}, \frac{2}{3})\}$;
(c) $E_0 = \{\alpha(\frac{1}{2}, 1) + (1-\alpha)(\frac{2}{3}, \frac{2}{3}); \alpha \in [0,1]\}$.

Démonstration. — Nous donnons seulement l'idée de la preuve en suivant Thuijsman au chapitre 12 dans Neyman et Sorin (2003). Pour une preuve complète veuillez consulter Sorin (1986).

Étape 1. — Nous prouvons (a), (b) se prouve de la même manière.

Remarquons que tout paiement d'équilibre doit donner au moins $\frac{1}{2}$ au joueur 1 et doit donner au moins $\frac{2}{3}$ au joueur 2. En effet, chaque joueur peut garantir au moins son montant (le joueur 2 en jouant

$\frac{1}{3}G + \frac{2}{3}D$ à chaque étape et le joueur 1 en jouant comme dans le Big-Match escompté $\frac{\lambda}{1+\lambda}H + \frac{1}{1+\lambda}B$).

Soit w le paiement d'équilibre le plus favorable au joueur 2 dans E_λ. Nous allons prouver que $w \leqslant \frac{2}{3}$ (et donc est égal à $\frac{2}{3}$).

Soit $\left(\sigma^1, \sigma^2\right)$ un équilibre dans Γ_λ tel que $g_\lambda^2\left(\sigma^1, \sigma^2\right) = w$. Soit w_1 (resp. w_2) le paiement de continuation sous $\left(\sigma^1, \sigma^2\right)$ si à l'étape 1 la paire (H, G) a été jouée (resp. (H, D) a été jouée).

Remarquez que w_1 et w_2 doivent aussi correspondre à des paiements d'équilibres dans Γ_λ pour le joueur 2 (car sinon le joueur 2 pourrait dévier). Donc on a :

$$w_1 \leqslant w \text{ et } \ w_2 \leqslant w.$$

Soit maintenant p la probabilité avec laquelle le joueur 1 joue B à la première étape sous σ^1 et soit q la probabilité avec laquelle le joueur 2 joue D sous σ^2. On vérifie facilement que si $\left(\sigma^1, \sigma^2\right)$ est un équilibre alors p et q sont dans $]0,1[$. Nous en déduisons alors que le joueur 2 est indifférent entre ces deux actions à la première étape, d'où

$$\begin{aligned} w &= 2p + (1-p)(1-\lambda)w_1 \\ &= 0p + (1-p)(\lambda + (1-\lambda)w_2) \end{aligned}$$

La ligne 2 (resp. 3) correspond à l'espérance de gain si G est jouée (resp. si D est jouée). Puisque $w_1 \leqslant w$ et $w_2 \leqslant w$ nous obtenons :

$$\begin{aligned} w &\leqslant 2p + (1-p)(1-\lambda)w \\ w &\leqslant (1-p)(\lambda + (1-\lambda)w). \end{aligned}$$

La deuxième inégalité devient

$$2 - w \geqslant (1-p)(2 - (1-\lambda)w),$$

et en utilisant la première inégalité nous trouvons

$$(2-w)\,(1-p)(\lambda + (1-\lambda)w) \geqslant (1-p)(2-(1-\lambda)w)w$$

qui est équivalente à $w \leqslant \frac{2}{3}$.

Un calcul similaire montre que le paiement maximal possible pour le joueur 1 n'excède pas $\frac{1}{2}$.

Étape 2. — Pour prouver que $E_0 \subset \{\alpha(\frac{1}{2},1)+(1-\alpha)(\frac{2}{3},\frac{2}{3}) \mid \alpha \in [0,1]\}$ Sorin (1986) donne l'argument suivant :

L'idée de la preuve est très simple : si la probabilité d'obtenir un paiement absorbant sur le chemin d'équilibre est moins que 1, alors après un certain temps le joueur 1 joue essentiellement l'action H; les paiements de continuations correspondants à partir de cette étape ne sont pas individuellement rationnels, d'où la contradiction.

Le lecteur est prié de consulter Sorin (1986) pour la traduction mathématique de cet argument.

Étape 3. — Nous prouvons que $\{\alpha(\frac{1}{2},1)+(1-\alpha)(\frac{2}{3},\frac{2}{3}) \mid \alpha \in [0,1]\} \subset E_0$.

L'argument se comprend facilement à l'aide d'un exemple. Prenons par exemple le couple de paiements $\left(\frac{7}{12},\frac{10}{12}\right)$ et expliquons la méthode de Sorin pour construire un ε-équilibre uniforme. Considérons le jeu auxiliaire à somme nulle suivant :

	G	D
H	$\frac{7}{12}$	$-\frac{5}{12}$
B	$-\frac{7}{12}{}^*$	$\frac{5}{12}{}^*$

Soit σ^1_ε une stratégie uniformément ε-optimale pour le joueur 1 dans le jeu auxiliaire, similaire à celle de Blackwell et Ferguson 1968. Soit x^2 la stratégie stationnaire du joueur 2 qui joue $\frac{5}{12}G+\frac{7}{12}D$ à chaque étape. Sous $(\sigma^1_\varepsilon, x^2)$ le jeu est absorbé avec probabilité 1 et induit donc un vecteur de paiements dans le vrai jeu exactement égal à $\left(\frac{7}{12},\frac{10}{12}\right)$. Donc, σ^1_ε est une meilleure réponse du joueur 1 dans le Paris-Match face à x^2 car sans absorption (en jouant H) le joueur 1 gagne seulement $\frac{5}{12}$. En utilisant la définition de la stratégie du joueur 1, il est possible de montrer que x^2 est uniformément ε-optimale face à σ^1_ε dans le Paris Match (voir Sorin (1986)). $\square$

13. Extensions

La construction d'une stratégie maxmin uniforme en cas d'observation imparfaite des actions a été faite d'une manière indépendante par Coulomb (2003) et Rosenberg, Solan et Vieille (2003).

Vieille (2000 a,b) a généralisé remarquablement le résultat de Mertens et Neyman (1982) aux jeux à deux joueurs et à somme non nulle en prouvant l'existence d'un paiement d'équilibre uniforme. Sa preuve, en deux étapes, consiste à réduire d'abord le problème d'existence à la classe des jeux récursifs (jeux dans lesquels les joueurs ont une fonction de paiement nulle dans tous les états non absorbants). Les jeux d'arrêts présentent la particularité d'être à la fois récursifs et absorbants d'où leur intérêt.

Exception faite de quelques cas particuliers, comme les jeux absorbants à trois joueurs résolus positivement par Solan (1999) qui prouve l'existence d'un paiement d'équilibre uniforme, le problème pour les jeux à plus de deux joueurs reste entièrement ouvert. Solan et Vieille (2003) ont exhibé un jeu d'arrêt à quatre joueurs qui prouve que pour montrer l'existence d'un paiement d'équilibre uniforme, il faut nécessairement inventer une nouvelle classe de stratégies, différentes de celles utilisées habituellement (comme celle dans la preuve de Mertens-Neyman 1981).

Plusieurs personnes conjecturent la non-existence d'un paiement d'équilibre uniforme pour tout jeu stochastique à n joueurs. Pour prouver cela, il y a deux possibilités : trouver un contre exemple ou prouver que l'existence induit un paradoxe mathématique. La non-existence a été prouvée en temps continu. Laraki, Solan et Vieille (2004) ont exhibé un jeu d'arrêt simple à trois joueurs qui n'admet pas d'équilibre approché, que ça soit au sens faible (escompté) ou fort (uniforme). Le même exemple en temps discret admet un équilibre stationnaire pour le jeu escompté (par Fink (1964)) et un paiement d'équilibre uniforme (par Solan (1999)).

La classe la plus simple à étudier est certainement celle des jeux d'arrêts. Dans un jeu d'arrêt, chaque joueur a le choix entre A (arrêter) ou C (continuer). Tant que le jeu n'est pas arrêté, le gain d'étape est zéro pour chaque joueur. La première fois qu'un ou plusieurs joueurs choisissent l'action A le jeu est absorbé. Le paiement d'absorption de chaque joueur dépend de la coalition des joueurs S qui ont arrêté le jour de l'absorption. À part une sous classe résolue positivement par Solan et Vieille (2001), le problème de l'existence d'un paiement d'équilibre uniforme reste entièrement ouvert (peut être pourriez vous le résoudre ?).

L'existence de la valeur dans les jeux escomptés a été généralisée à des espaces d'états et d'actions plus généraux par Maitra et Parthasarathy (1970) et d'autres auteurs jusqu'au résultat très général de Nowak (1985).

Une façon très naturelle d'étudier v_λ et v_T consiste à étudier les propriétés mathématiques de l'opérateur de Shapley et de sa dérivée pour en déduire la convergence et une caractérisation de la limite. Cette approche a été utilisée pour la première fois par Kohlberg (1974) dans le cadre restreint des jeux absorbants et a été définie et étudiée d'une manière générale pour les jeux stochastiques à somme nulle par Rosenberg et Sorin (2001). Elle a été appliquée par les deux auteurs pour étudier les jeux absorbants avec des espaces d'actions compacts mais aussi pour l'étude des jeux répétés à information incomplète des deux côtés (*cf.* texte sur les jeux répétés (ce volume)). La même approche permet à Rosenberg (2000) de montrer l'existence et la caractérisation variationnelle de la valeur asymptotique des jeux absorbants et information incomplète d'un côté.

Nous invitons les lecteurs intéressés par d'autres résultats ou extensions à consulter le livre sur les jeux stochastiques édités par Neyman et Sorin (2003), le livre de Sorin (2002) ou encore l'ouvrage de référence sur les jeux répétés par Mertens Sorin et Zamir (1994). Pour une étude détaillée de l'approche infinie (avec des fonctions de paiements mesurables définies sur l'ensemble des parties comme les liminf et limsup des sommes de Césaro mais aussi d'autres objectifs comme atteindre ou éviter un sous ensemble mesurable de H_∞) vous pouvez consulter l'ouvrage de référence de Maitra et Sudderth (1996).

Bibliographie

Blackwell (D.)

[1962] Discrete dynamic programming, *Annals of Mathematical Statistics*, 33 (1962), p. 719–726.

Blackwell (D.) & Ferguson (T.)

[1968] The Big Match, *Annals of Mathematical Statistics*, 33 (1968), p. 882–886.

Benedetti (R.) & Risler (J.-J.)

[1990] *Real algebraic and semi-algebraic sets*, Paris : Hermann, 1990.

Bewley (T.) & Kohlberg (E.)
[1976a] The asymptotic theory of stochastic games, *Mathematics of Operation Research*, 1 (1976), p. 197–208.
[1976b] The asymptotic solution of a recursion equation occuring in stochastic games, *Mathematics of Operation Research*, 1 (1976), p. 321–336.

Coulomb (J.M.)
[1992] Repeated games with absorbing states and no signals, *International Journal of Game Theory*, 21 (1992), p. 161–174.
[2001] Repeated games with absorbing states and signaling structure, *Mathematics of Operation Research*, 26 (2001), p. 286–303.
[2003] Stochastic games without perfect monitoring, *International Journal of Game Theory*, 32 (2003), p. 723–96.

Fink (A.M.)
[1964] Equilibrium in a stochastic n-person game, *J. Sci. Hiroshima Univ.*, 28 (1964), p. 89–93.

Foster (O.)
[1981] *Lectures on Riemann surfaces*, Springer, 1981.

Kohlberg (E.)
[1974] Repeated games with absorbing states, *Annals of Statistics*, 2 (1974), p. 724–738.

Laraki (R.)
[2001] Variational inequalities, system of functional equations, and incomplete information repeated games, *SIAM Journal of Control and Optimization*, 40(2) (2001), p. 516–524.
[2002] Repeated games with lack information on one side : the dual differential approach, *Mathematics of Operation Research*, 27(2) (2002), p. 419–440.

Laraki (R.), Solan (E.) & Vieille (N.)
[2005] Continuous-time games of timing, *Journal of Economic Theory*, 120 (2005), p. 206–238.

Laraki (R.)
[2006] A variational approach in zero-sum absorbing games, 2006 ; preprint.

Maitra (A.) & Parthasarathy (T.)
[1970] On stochastic games, *Journal of Optimization Theory and Applications*, 5 (1970), p. 289–300.

Maitra (A.) & Sudderth (W.)
[1996] *Discrete gambling and stochastic games*, Springer, 1996.

Mertens (J.-F.) & Neyman (A.)
[1981] Stochastic games, *International Journal of Game Theory*, 10 (1981), p. 53–66.

Mertens (J.-F.), Sorin (S.) & Zamir (S.)
[1994] Repeated games, CORE D.P., 1994, p. 9420–21–22.

Neyman (A.) & Sorin (S.)
[2003] *Stochastic games and applications*, NATO Science Series, Kluwer Academic Publishers, 2003.
Nowak (A.S.)
[1985] Universally measurable strategies in zero-sum stochastic games, *Annals of Probability*, 13 (1985), p. 269–287.
Rosenberg (D.)
[2000] Zero-sum absorbing games with incomplete information on one side : asymptotic analysis, *SIAM Journal of Control and Optimization*, 39 (2000), p. 208–225.
Rosenberg (D.) & Sorin (S.)
[2001] An operator approach to zero-sum repeated games, *Israël Journal of Mathematics*, 121 (2001), p. 221–246.
Rosenberg (D.), Solan (E.) & Vieille. (N.)
[2002a] Blackwell Optimality in Markov Decision Processes with Partial Observation, *Annals of Statistics*, 30 (2002), p. 1178–1193.
[2002b] The MaxMin value of stochastic games with imperfect monitoring, *International Journal of Game Theory*, 32 (2002), p. 133–150.
Solan (E.)
[1999] Three-Player absorbing games, *Mathematics of Operation Research*, 24 (1999), p. 669–698.
Solan (E.) & Vieille (N.)
[2001] Quitting games, *Mathematics of Operation Research*, 26 (2001), p. 265–285.
[2003] Quitting games – an example, *International Journal of Game Theory*, 31 (2003), p. 365–381.
Solan (E.)
[2006] Stochastic games, 2006 ; a non-published master course.
Sorin (S.)
[1986] Asymptotic properties of a non zero-sum stochastic game, *International Journal of Game Theory*, 15 (1986), p. 101–107.
[2002] *A First course on zero-sum repeated games*, Springer-Verlag, 2002.
Shapley (L.S.)
[1953] Stochastic games, *Proceedings of the National Academy of Sciences of the U. S. A.*, 39 (1953), p. 1095–1100.
Takahashi (M.)
[1964] Equilibrium points of stochastic non-cooperative n-person game, *J. Sci. Hiroshima Univ.*, 28 (1964), p. 95–99.
Vieille (N.)
[2000a] Two-player stochastic games I : a reduction, *Israël Journal of Mathematics*, 119 (2000), p. 55–91.

[2000b] Two-player stochastic games II : the case of recursive games, *Israël Journal of Mathematics*, 119 (2000), p. 93–126.

R. LARAKI, CNRS, Laboratoire d'Économétrie, École polytechnique, 1 rue Descartes, 75005 Paris • *E-mail :* `Rida.Laraki@shs.polytechnique.fr`
Url : `http://ceco.polytechnique.fr/home/laraki/FR`